Philipps-Universität Marburg

Asymptotics for selected Risk Measures under general assumptions

Dissertation

zur Erlangung des Doktorgrades
der Mathematisch- Naturwissenschaftlichen Fakultäten
der Philipps-Universität Marburg

vorgelegt von

Tobias Zwingmann

Master of Science

aus Kassel

Marburg, 2018

Betreuer: Prof. Dr. Hajo Holzmann
Zweitgutachter: Prof. Dr. Axel Bücher
Eingereicht: 28. November 2018
Tag der Disputation: 11. März 2019
Erscheinungsort und -jahr: Marburg 2019
Hochschulkennziffer: 1180

To Fine and David.

Acknowledgements

I am grateful for the support of many people during my doctoral studies. Hajo, thanks for making the present work possible in the first place by proposing the problem and suggesting methods; thank you for your encouragement and listening, not only but also to ideas; and – last but not least – thank you for your patience and consideration of my family situation.

I want to thank Professor Axel Bücher for accepting the Koreferat.

The thesis would have not been written yet without you, Rini. Thank you for giving me the time I needed and silence if needed. Thanks for your interest in the studies and support of my person.

Sara and Jonas, thank you, thank you, thank you for smiling, and diverting, and explaining the world to me.

I would like to thank my colleagues at the Department of Mathematics, especially Viktor Bengs and Heiko Werner for fruitful discussions and (playful) diversion. Viktor, I think `R` and me are friends because of you at last – except `rgl`.

In addition, I thank Miri, Luki and Hannah for helpful corrections when reading the thesis. Miri, I hope you know the comma rules more precisely now.

Further, I gratefully acknowledge the financial support of the Cusanuswerk. Thanks for many opportunities of scientific and personal exchange.

Contents

List of Figures

List of Tables

List of Symbols

Functions

$\mathbb{1}(x \in B)$	Indicator function.
$\|\cdot\|_D$	Supremum over the set D.
$\mathcal{C}(K)$	Set of continuous functions $K \to \mathbb{R}$.
$\mathcal{D}([0,1])$	Set of càdlàg-functions $[0,1] \to \mathbb{R}$.
$\partial^-(h)(t)$, $\partial^+(h)(t)$	Left- and right-sided derivative of h in t.
$g(F)$	The vector $\left(\int_{\mathbb{R}} g^i(y)\,\mathrm{d}F(y)\right)_{1\le i\le s}$ for a distribution function F.
$h(x-), h(x+)$	Left- and right-sided limit of h in x.
h^{Inv}	(Pseudo-) Inverse of a function h.
$h_\wedge$, $h_\vee$	Lower- and upper-semicontinuous hulls of h.
$\mathrm{id}_{\mathcal{X}}$	Identity function on $\mathcal{X}$.
$J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})$	Entropy with bracketing of the set $\mathcal{H}$.
$\ell^\infty(K)$	Set of bounded functions $K \to \mathbb{R}$.
$\mathcal{L}_\infty$	Set of bounded random variables.
$\mathcal{L}_p$	Set of random variables with finite p-th moment.
$N_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})$	Bracketing number of the set $\mathcal{H}$.
$x \wedge y$, $x \vee y$	Minimum/maximum of $x, y \in \mathbb{R}$.
$\lceil x \rceil$	Ceiling function; smallest natural number greater than or equal to x.

(Semi-)metrics

$\|\cdot\|$	Supremum norm.

d_e	Euclidean metric.
$\mathrm{d}_{\mathbb{F}}$	Metric generating the Fell-Matheron-topology.
d_{hypi}	Hypi-metric.
$\mathrm{d}_{s,1}$	M_1-metric.
$\mathrm{d}_{s,2}$	M_2-semimetric.

Probability

$(\Omega, \mathcal{A}, \mathrm{P})$	Probability space.
$\rightsquigarrow$	Convergence in distribution for random processes.
$\xrightarrow{\mathcal{L}}$	Convergence in distribution for random variables.
$\xrightarrow{\mathrm{P}}$	Convergence in probability.
Cov, Var	(Co-)Variance.
δ_x	Dirac-measure with mass 1 in x.
$\mathbb{E}_n[\,]$, $\mathbb{E}[\,]$	(Empirical) expectation.
$\mathbb{E}^\circ$, P°, $\mathbb{E}_\circ$, $\mathrm{P}_\circ$	Outer/inner expectation and probability.
F_n, F	(Empirical) distribution function.
$\mathcal{N}(\mu, \Sigma)$	(Multivariate) normal distribution with mean μ and covariance matrix Σ.
$O_{\mathrm{P}}(a_n)$, $o_{\mathrm{P}}(a_n)$	Stochastic Landau symbols.
$\mathcal{U}(a,b)$	Uniform distribution on (a,b).
$Y_{i:n}$	The i-th order statistic.

Risk Measures

es_α	Expected Shortfall at level α.
κ_m	Spectral risk measure associated to the probability measure m.
μ_τ	Expectile at level τ.
$q_\alpha^\pm, q_\alpha$	Upper, lower and any α-quantile.

Topology

$B_\varepsilon^{\mathcal{X}}(x)$	Open ball of radius ε around x in $\mathcal{X}$.
$B^{\delta,\mathcal{X}}$	The δ-enlargment of a set B in $\mathcal{X}$.
$\mathrm{cl}(B)$	Closure of a set B.
∂B	Boundary of a set B.
$\mathrm{int}(B)$	Interior of a set B.

Introduction

The first questions when reading the title could be: What is risk and how can we measure it, especially in practice? It is widely accepted that when considering a real world mechanism, assertions about its future state have to be of probabilistic nature. Thus, there is the potential of deviations from the expected outcome of the mechanism, which we call risk. With the aid of risk measures these deviations can be quantified, for example enabling decision-making grounded on this quantification. In practice we have to estimate the desired values based on observations of the considered mechanism.

One question we address in this work is how certain estimates for risk measures behave in statistical terms. More precisely, we choose three useful risk measures and estimates thereof, for which we prove (functional) central limit theorems in non-standard situations, laying the base for further statistical examination.

Exemplary considering applications in financial markets, the properties of a chosen risk measure should reflect agreed principles of risk. First, if we have no asset, there should be no risk, meaning the risk measure should assign 0 to that asset. This is called *normalization*. Second, if we know that an asset has a guaranteed return, adding this to our portfolio should decrease the risk by the secure return, which is called *translation invariance*. Third, if there is an asset, which always yields better returns than another, the first ought to have a higher risk. We call this *monotonicity* of the risk measure.

The properties so far do not capture one of the most important principles in economics, namely the diversification principle. By this, the risk of two assets together should not exceed the sum of the risk of the individual ones. Mathematically this is called *sub-additivity* of the risk measure at hand. Additionally, if we buy another share of the asset, the risk should scale according to this proportion. Then the risk measure is *positive homogeneous*. A risk measure having all of these five properties is called *coherent*. Coherence was first introduced by Artzner et al. (1999), where the above properties are further justified. Some works even decide to use the term risk measure only for coherent risk measures (Acerbi and Tasche, 2002a, page 3). A last property we want to mention here is the *comonotonicity* of a risk measure. Assume we have two positions, which always evolve in the same direction, meaning that one position gains value if and only if the other does (not necessarily the same amount) and similarly for loosing value. Then we should not be able to exploit the diversification principle, as the evolution of the two is too alike. This implies that the risk of two such positions added equals the sum of the individual risk measures.

Perhaps one of the most common risk measures – besides the mean and the variance – is the so called *Value at Risk*; historically it has probably been the first in wide use. This measure returns a value, beyond which losses are only suffered with a fixed probability; the latter is called the (confidence) level of the Value at Risk. From the mathematical point of view, the Value at Risk is a certain quantile of the profit-and-loss distribution.

The importance of the Value at Risk stems among others from the fact that the Basel II and III frameworks explicitly incorporate it and give regulations for calculating Value at Risk-models within banks (see Basel, 2017). In addition, the Value at Risk can mathematically be defined as the unique minimum of a deterministic function – it is *elicitable* –, which opens the path to many statistical tools, as regression frameworks (see Koenker, 2005) and comparative backtesting; additionally, this fact is important for us in the course of the thesis. The function to be minimized is called *scoring function* in general; for the Value at Risk the aptronym "check-function" appears in the literature.

On the other hand, there is a major drawback to be named for the Value at Risk. In particular, that risk measure is not coherent, as it is not always sub-additive and therefore could discourage diversification. This leads us to the second measure investigated through the thesis, the *Expected Shortfall*, which the Basel Committee of Banking Supervision also recommends to use (see Basel, 2013, page 3).

The Expected Shortfall at a chosen level is the average of all Value at Risks pending that fixed level. In some cases, for example for continuous distributions, it equals the expected loss of the profit-and-loss distribution, given that the loss is higher than the Value at Risk at the level.

It turns out that this risk measure is coherent (Acerbi and Tasche, 2002a), but unfortunately not accessible by minimizing a scoring function as in the quantile case (Weber, 2006; Gneiting and Raftery, 2007). Thus, especially comparative backtesting of the Expected Shortfall is questionable (Gneiting, 2011a). The good news is that the situation changes when considering Value at Risk and Expected Shortfall simultaneously, as, recently, a scoring function for the bivariate risk measure (Value at Risk, Expected Shortfall) was constructed by Fissler and Ziegel (2016). A major part of the thesis works with this pair of risk measures and generalizations thereof, using the respective scoring function to deduce a central limit theorem for the empirical versions of the risk measures. For the pair (Value at Risk, Expected Shortfall) this is done under weak conditions on the first entry, especially dropping the standard assumption of an existing and strictly positive derivative of the underlying distribution function in the Value at Risk.

The generalizations of (Value at Risk, Expected Shortfall) considered in the present work are twofold: First, the Expected Shortfall is a special case of so called *spectral risk measures*, second, it can be seen as a Bayes risk; both paths are detailed in the course of the thesis.

If we want to comprise the benefits of the Value at Risk and the Expected Shortfall, we are directly led to the *expectile* at some confidence level to be chosen. This risk measure is on the one hand identifiable as the unique minimum of a scoring function and on the other hand coherent. This is actually a unique property among risk measures (Ziegel, 2016).

The expectile is defined as the expectation of the distribution at hand, conditional on being below or above some threshold where deviations up- and downwards are weighted differently. So while the Value at Risk does not take the height of any loss above some threshold into account and the Expected Shortfall only considers high losses, the expectile takes both high and low losses into account, giving them different importance. In practice it is not directly obvious why a risk measure should account for both high and low outcomes, which is the major criticism for the expectile. But this can be justified in financial terms by regarding high profits as subject to tax, whereas interpreting high losses as a tax shield; see Ehm et al. (2016).

In the thesis we consider the expectile at several levels simultaneously, showing a functional central limit theorem for the empirical estimate of the expectile curve under weak assumptions on the underlying profit-and-loss distribution.

The thesis is structured as follows. To start with, in Chapter 1 we rigorously introduce the mathematical objects and basic notation needed throughout the work. Additionally, we present a first result about scoring functions concerning the decomposition in elementary scores and specify a result from Frongillo and Kash (2015a) related to the elicitability of Bayes risks. After reviewing the concepts of M-estimation, Chapter 1 also comprises an assertion, which yields the possibility to deal with multi-dimensional M-estimates, where the entries can have different rates of convergence. Chapter 2 revisits the Hoffmann-Jørgensen theory of weak convergence in metric spaces and shows how to extend this to semimetric spaces. The latter is applied to the space of bounded functions equipped with the hypi-semimetric of Bücher et al. (2014). Thereafter we prove new results about hypi-convergence, establishing a general scheme to show weak convergence results for transformed processes if the transformation is semi-Hadamard differentiable. Concluding Chapter 2 we take up the issue of Skorohod convergence, which gives another interpretation of the hypi-distance. Chapter 3 investigates the behaviour of empirical estimators for the pair (Value at Risk, Expected Shortfall) concerning a central limit theorem by using the scoring function of Fissler and Ziegel (2016). We show that the rate of convergence of the estimator for the Value at Risk does not influence the rate of convergence of the empirical Expected Shortfall. The chapter also includes multivariate versions of the results, precisely asymptotic considerations for a vector of k empirical Value at Risks and respective Expected Shortfalls, which facilitates the examination of the weak convergence of empirical spectral risk measures. The scheme developed in Chapter 3 is extended to general Bayes risks in Chapter 4. Chapter 5 is dedicated to the expectile – viewed as a random function on a closed interval – giving a weak convergence result for the empirical expectile process with respect to the hypi-semimetric under general assumptions. The framework used there is also applied to quantiles, generalizing well-known results about the asymptotic behaviour of the empirical quantile process.

Chapter 1

Scores and M-Estimation

In this chapter we rigorously introduced the concepts needed in the work and set the basic notation used. In addition, we touched on the risk measures discussed throughout the thesis, giving background material and connections. Proceeding to the concept of M-estimation, we gave a result enabling us to deal with multi-dimensional M-estimates with entries having different rates of convergence.

1.1 Introduction

Imagine we as experts are asked to give an opinion about the risk of some uncertain position. Certainly we want to get paid for our answer, such that the enquirer faces a more or less severe problem. How can he incentivize us with his payment to issue true information? A way to achieve this would be letting the payment depend on your answer and the outcome of the event.

Assume that we are therefore given a function $S(x; y)$, which describes our gain when issuing x, while outcome y occurs. Which value x do we choose when knowing the outcome y? Unfortunately, we do not know the value of y and thus consider it as a random variable, writing Y instead. Supposing we have a clue about the distribution of Y, we can calculate our expected payment $\mathbb{E}[S(x; Y)]$ for any x and try to maximize this. Giving that maximizer as answer to our customer, "on average" we make the best decision possible – at least with respect to our wallet.

Functions S of this type are called *consistent scoring functions* or *consistent scores*. A parameter for which a *strictly consistent* score exists, which means that the expected score is uniquely maximized at that parameter, is called *elicitable* – a concept developed around Savage (1971) and Osband (1985) and heavily investigated thereafter, see for example Gneiting (2011a) and references therein. Usage of scores can be found in a variety of scientific fields, such as statistics (Savage, 1971; Osband, 1985; Gneiting and

Raftery, 2007), machine learning (Steinwart et al., 2014; Frongillo and Kash, 2015b), economics (Lambert and Shoham, 2009; Lambert, 2013) as well as finance (Emmer et al., 2015).

Taking a Bayesian point of view, scores are often referred to as utilities. Then, the above strategy for the choice of parameters represents the idea in Bayesian decision theory of maximizing the expected utility of a posterior distribution (Lehmann and Casella, 2006). Scoring functions can also be seen as a generalization of so called *scoring rules*, which aim for reporting the whole distribution of the random variable Y; see Frongillo and Kash (2015b).

In financial mathematics, elicitability is considered an essential property of parameters, as they allow for comparative backtesting (Nolde and Ziegel, 2017), regression frameworks as quantile or expectile regression (see Koenker (2005) and Newey and McFadden (1994) respectively) and M-estimation considered below. Thus, especially risk measures used in practice should be elicitable in order to access these benefits.

But not every parameter is elicitable, for example the variance or the mode of a random variable (Heinrich, 2014) are not. A main part of research about scoring functions therefore is the question which classes of parameters are elicitable while characterizing the set of appropriate scoring functions for the parameter at hand. Results can be found in Osband (1985) and Lambert, Pennock, et al. (2008); see also Lambert (2013), who links elicitability of a parameter to the convexity of so called level sets, saying that a parameter, which can be expressed as linear constraint of the distribution, is elicitable. This immediately rules out the variance.

In many cases the problem of non-elicitability can be circumvented. It turns out, that many non-elicitable parameters are one entry of an elicitable vector of k parameters. For example the vector (mean, variance) is elicitable, similar for (Value at Risk, Expected Shortfall) as indicated above; for recent results we refer to A. Agarwal and S. Agarwal (2015), Frongillo and Kash (2015b) and Fissler and Ziegel (2016).

Until now we assumed more or less implicitly, that we have access to the true distribution of Y, such that we can calculate the expected score $\mathbb{E}\left[S(x;Y)\right]$ to be maximized. In practice we often only know a data dependent approximation for the true expected score and maximizing this in general does not yield the true parameter but an estimate thereof. Such estimators are called *M-estimators*. Introduced in the 60s by Huber (1964) and Huber (1967), they were a natural generalization of the maximum likelihood estimation and played an important role for robust statistics. A comprehensive treatment for the asymptotic behaviour, such as consistency and asymptotic normality, can be found in van der Vaart (1998).

Regarding the estimation of an elicitable parameter leads to another question currently addressed in the literature. Asking two experts about a forecast or estimation of the parameter of interest – leaving open how to obtain them –, we probably get at least two different answers. Which one is preferable to the other? A possible way in deciding this is to look whether the expected score at the one issued value is bigger than the expected score at the other one *for any possible score*. This reflects the idea, that the value of the score, $S(x; y)$, can be interpreted as the economic loss when choosing x while y occurs. Several testing frameworks have been proposed in this context (Diebold and Mariano, 1995; Clark and McCracken, 2001; Ziegel et al., 2017). We contribute to this field by proving a result for Bayes risks about the decomposition of scores in elementary scores. The latter are used to construct *Murphy diagrams*, allowing for visual comparison of the performance of the estimates. See Ehm et al. (2016) for quantiles and expectiles and Ziegel et al. (2017) for the Expected Shortfall.

The course of this chapter is as follows. We start with introducing some basic objects and notation in Section 1.2, followed by an introduction to risk measures in Section 1.3. In Section 1.4 we turn to elicitability and (elementary) scores and refine a result about the elicitability of Bayes risks given in Frongillo and Kash (2015a). A first result of this thesis about the decomposition of scores for the Bayes risk into elementary scores is presented as well. Section 1.5 is devoted to M-estimation, where we especially consider a consistency result for multi-dimensional estimates, in which the entries can have different rates of convergence. All proofs are deferred to Section 1.6.

1.2 Basic notation

We always work over a complete probability space $(\Omega, \mathcal{A}, \mathrm{P})$. We write $P \otimes P$ for the product measure. Elements in Ω are denoted by ω. Random variables, meaning Borel-measurable maps $\Omega \to \overline{\mathbb{R}}$, are depicted with upper case letters Y; all random variables are collected in the set $\mathcal{L}$. For a random variable Y and a set $A \subset \overline{\mathbb{R}}$ we write $\{Y \in A\}$ as usual for the preimage of A under Y, precisely we set $\{Y \in A\} = \{\omega \in \Omega \,|\, Y(\omega) \in A\}$. Notations like $\{Y \leq y\}$ are interpreted accordingly.

If $\mathcal{X}$ is any set and $B \subset \mathcal{X}$, we write $\mathbb{1}(\cdot \in B)$ for the indicator function, namely $\mathbb{1}(x \in B) = 1$ if $x \in B$ and $\mathbb{1}(x \in B) = 0$ otherwise. The function $\mathrm{id}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}$, $x \mapsto x$, is the identity on $\mathcal{X}$. The cardinality of the set B is abbreviated with $|B|$. In the case that $\mathcal{X}$ is a topological space, $\mathrm{int}(B)$, $\mathrm{cl}(B)$ and ∂B are the interior, closure and boundary of the set B, respectively. For a metric space $(\mathcal{X}, d)$ we denote with $B_\varepsilon^{\mathcal{X}}(x) = \{y \in \mathcal{X} \,|\, d(x, y) < \varepsilon\}$ the ε-ball around $x \in \mathcal{X}$. The δ-enlargement of the set B is $B^{\delta,\mathcal{X}} = \bigcup_{x \in B} B_\delta^{\mathcal{X}}(x)$. We sometimes suppress the index $\mathcal{X}$.

For a random variable Y with distribution function F we write $Y \sim F$; when having a sequence $Y_1, \ldots, Y_n$ of random variables. The map δ_x is the Dirac measure, putting unit mass in the point x. With $Y_{1:n} \leq \ldots \leq Y_{n:n}$ we denote the order statistics of the present sample. If two random variables Y and Z have the same distribution function, we write $Y \stackrel{\mathrm{d}}{=} Z$.

We use $\mathbb{E}_F[f(Y)]$ for notating the expectation of a (measurable, integrable) transformation f of the random variable Y with distribution function F, precisely we set $\mathbb{E}_F[f(Y)] = \int_{\mathbb{R}} f(y) \,\mathrm{d}F(y) = \int_\Omega f(Y) \,\mathrm{dP}$; sometimes we drop the subscript and only write

$\mathbb{E}[f(Y)]$. Similarly, we use $\mathbb{E}_n[f(Y)] = \int_{\mathbb{R}} f(y)\,\mathrm{d}F_n(y)$ for the empirical expectation and write $\mathbb{G}_n[f(Y)] = \sqrt{n}(\mathbb{E}_n - \mathbb{E})[f(Y)]$ for the centred and rescaled version thereof. The variance of a random variable $Y \sim F$ with finite second moment is denoted with $\mathrm{Var}_F(Y)$ or $\mathrm{Var}(Y)$.

The set of random variables Y with finite p-th moment, $\mathbb{E}[|Y|^p] < \infty$ for $p \in (0, \infty)$, is abbreviated with $\mathcal{L}_p = \mathcal{L}_p(\Omega, \mathcal{A}, \mathrm{P})$; in $\mathcal{L}_\infty = \mathcal{L}_\infty(\Omega, \mathcal{A}, \mathrm{P})$ we collect all bounded random variables.

We write $\mathcal{N}(\mu, \sigma^2)$ for a univariate normal distribution with mean μ and variance σ^2, the corresponding distribution function is denoted with Φ_{μ,σ^2}. Multivariate normal distributions are indicated with $\mathcal{N}(\mu, \Sigma)$, where now μ is the vector of means and Σ the covariance matrix. Next, $\mathcal{U}(a, b)$ abbreviates the uniform distribution on (a, b), $a < b$.

Let $\mathcal{F}$ be a family of distribution functions on $\mathbb{R}$ and $g = (g^1, \ldots, g^k) : \mathbb{R} \to \mathbb{R}^k$ be a function. If every g^s is F-integrable for any $F \in \mathcal{F}$, we say that g is $\mathcal{F}$-integrable. In that case, we just write $g(F)$ for the vector $\big(\int_{\mathbb{R}} g^s(y)\,\mathrm{d}F(y)\big)_{1 \le s \le k}$.

For considering asymptotic behaviour we need several types of convergence for a sequence of random variables $(Y_n)_n$ to Y. We use $\xrightarrow{\mathrm{P}}$ for indicating convergence in probability with respect to P. The arrow $\xrightarrow{\mathcal{L}}$ denotes weak convergence of random variables, whereas $\rightsquigarrow$ is used for weak convergence of stochastic processes (see Chapter 2). The sequence $(Y_n)_n$ is stochastically bounded, denoted with $O_{\mathrm{P}}(1)$, if for any $\varepsilon > 0$ there is a constant $C > 0$, such that $\sup_n \mathrm{P}(|Y_n| \ge C) \le \varepsilon$. If ${}^{Y_n}/_{a_n}$ is stochastically bounded for a sequence a_n, we write $Y_n = O_{\mathrm{P}}(a_n)$; if ${}^{Y_n}/_{a_n}$ converges to zero in probability, we write $Y_n = o_{\mathrm{P}}(a_n)$. We use $O_{\mathrm{P}}(a_n)$ and $o_{\mathrm{P}}(a_n)$ for sequences of processes as well.

Having a function h with existing left- and right-sided limits at a point x, we use $h(x-) = \lim_{y \nearrow x} h(y)$ and $h(x+) = \lim_{y \searrow x} h(y)$ to denote these left- and right-sided limits in x respectively. In order not to confuse fractions with the inverse of a function, we write $c^{-1} = {}^1/_c$ for fractions and use h^{Inv} for the (generalized) inverse of a function h. More precisely, if $h : \mathbb{R} \to \mathbb{R}$, we denote with h^{Inv} the pseudo-inverse of h, $h^{\mathrm{Inv}}(y) = \inf\{x \in \mathbb{R} |\, h(x) \ge y\}$. If however $h : D \to D'$ is an invertible function where the sets D and D' are arbitrary (mainly $\mathbb{R}$ or a space of functions), we write h^{Inv} for the inverse function of h. Note that if $h : \mathbb{R} \to \mathbb{R}$ is invertible, the pseudo-inverse reduces to the true inverse, such that these notations do not contradict each other. Additionally, if $h : D \to \mathbb{R}$, we set $\|h\|_D = \sup_{x \in D} |h(x)|$ as the supremum of h over D; this is often abbreviated with $\|h\|$, if D is clear from the context.

Given $x, y \in \mathbb{R}$, we set $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$. If $x \ge 0$, we denote with $\lceil x \rceil$ the smallest natural number greater than or equal to x.

1.3 Risk measures

In this section we define risk measures in general and give first concrete examples thereof, which are used in the later chapters. We indicate some connections but mostly refer to the literature for more extensive studies.

1.1 Definition.

A *risk measure* is a map $\rho : \mathcal{L} \to \mathbb{R} \cup \{\infty\}$, $Y \mapsto \rho(Y)$.
A risk measure ρ is said to be

i) *normalized* if $\rho(0) = 0$;

ii) *translation invariant* if for all $Y \in \mathcal{L}$ and $c \in \mathbb{R}$ it holds that $\rho(Y+c) = \rho(Y) - c$;

iii) *monotone* if for $Y_1, Y_2 \in \mathcal{L}$ with $Y_1 \leq Y_2$ almost surely $\rho(Y_1) \leq \rho(Y_2)$ is satisfied;

iv) *sub-additive* if for all $Y_1, Y_2 \in \mathcal{L}$, $\rho(Y_1 + Y_2) \leq \rho(Y_1) + \rho(Y_2)$ is valid;

v) *positively homogeneous*, provided for any $\lambda \geq 0$ and $Y \in \mathcal{L}$ it holds that $\rho(\lambda Y) = \lambda\, \rho(Y)$.

It is called *coherent* if properties i) – v) do hold.
Two variables $Y_1, Y_2 \in \mathcal{L}$ are *comonotone* if $\big(Y_1(\omega) - Y_1(\omega')\big)\big(Y_2(\omega) - Y_2(\omega')\big) \geq 0$ $\mathrm{P} \otimes \mathrm{P}$-almost surely.
The risk measure ρ is

vi) *comonotonic* if $\rho(Y_1 + Y_2) = \rho(Y_1) + \rho(Y_2)$ whenever Y_1 and Y_2 are comonotone;

vii) *law-invariant* if for any $Y_1, Y_2 \in \mathcal{L}$ the distributional equality $Y_1 \stackrel{\mathrm{d}}{=} Y_2$ implies $\rho(Y_1) = \rho(Y_2)$.

The map ρ is named *spectral* risk measure, if it is a comonotonic and coherent risk measure.

Having a law-invariant risk measure ρ and a set of distribution functions $\mathcal{F}$, we can consider ρ as a map $\mathcal{F} \to \mathbb{R} \cup \{\infty\}$ by writing $\rho(F) = \rho(Y)$ for any $Y \in \mathcal{L}$ with distribution function F, where the concrete choice of Y with distribution function F is rendered irrelevant by the law-invariance of ρ. The concrete risk measures dealt with later on are always law invariant, thus the latter interpretation of ρ is dominant in the following.

Note here that for connecting the above properties to the economical interpretation given in the introduction, we have to assume the distribution function F of Y to model the profit-and-loss distribution. Hence, negative and positive values of Y indicate loss and profit, respectively.

Now let us introduce some of the risk measures, which were mentioned in the introduction and are used in the course of the thesis. We start with the Value at Risk, for which we first need to define quantiles. After that we turn to the Expected Shortfall and a generalization thereof.

1.2 Definition (Quantile).

Let F be a distribution function and $\alpha \in [0,1]$. Then a value q is an *α-quantile* of F if $F(q) \geq \alpha \geq F(q-)$ holds.
A number q is an *α-quantile* of $Y \in \mathcal{L}$ if $\mathrm{P}(Y \leq q) \geq \alpha \geq \mathrm{P}(Y < q)$.

Let us briefly recall some properties of quantiles for a distribution function F.

1.3 Remark.

i) The set of α-quantiles of F for a fixed level $\alpha \in (0,1)$ is a closed interval. If there is no mass at infinity, there are finite values $q_\alpha^-(F)$ and $q_\alpha^+(F)$ such that this interval is given by $[q_\alpha^-(F),\, q_\alpha^+(F)]$. These points do satisfy $q_\alpha^+(F) = \inf\{y \,|\, F(y) > \alpha\}$ and $q_\alpha^-(F) = \inf\{y \,|\, F(y) \geq \alpha\} = F^{\mathrm{Inv}}(\alpha)$ and are called *lower* and *upper* α-quantile.

ii) Using the representation in i), the quantile functions $\alpha \mapsto q_\alpha^-(F)$, $\alpha \mapsto q_\alpha^+(F)$ are monotonically increasing.

iii) If F is strictly increasing in the α-quantile q_α, meaning $F(y) = \alpha$ for at most one y, then $q_\alpha^-(F) = q_\alpha = q_\alpha^+(F)$. This assertion is also necessary.

For further properties we refer to Embrechts (2013). ◇

The Value at Risk now is defined by giving one of the quantiles a fancy name.

1.4 Definition (Value at Risk).

Let $Y \in \mathcal{L}$ be distributed according to F_Y, then

$$VaR_\alpha(Y) = VaR_\alpha(F_Y) = -\inf\{y \in \mathbb{R} \,|\, F_Y(y) > \alpha\} = -q_\alpha^+(F_Y) \tag{1.1}$$

is the *Value at Risk at level* $\alpha \in [0,1]$.

The Value at Risk at level $\alpha \in [0,1]$ for a position gives the value below which no loss is attained with given probability α. Put the other way round, it is the value up to which $\alpha 100\%$ of possible losses do lie. The right hand side in (1.1) only depends on the distribution of Y, such that the Value at Risk is law-invariant. We can as well define the Value at Risk by using the distribution function F_{-Y} of $-Y$ as

$$\begin{aligned} -q_\alpha^+(F_Y) &= -\inf\{y \,|\, \mathrm{P}(Y \leq y) > \alpha\} = \sup\{-y \,|\, \mathrm{P}(-Y \geq -y) > \alpha\} \\ &= \sup\{x \,|\, \mathrm{P}(-Y < x) < 1-\alpha\} = \inf\{x \,|\, \mathrm{P}(-Y \leq x) \geq 1-\alpha\} \end{aligned}$$

and thus $VaR_\alpha(Y) = q^-_{1-\alpha}(F_{-Y})$. The Value at Risk is sub-additive in some situations (Danielsson et al., 2005), but not in general (Artzner et al., 1999; Acerbi, Nordio, et al., 2001). In addition, due to possible jumps in the distribution function F_Y, small changes in the confidence level α can result in severe changes of the Value at Risk – the map $\alpha \mapsto VaR_\alpha(F_Y)$ is not continuous.

The Value at Risk does not take into account the height of losses beyond a certain amount. This, however, is what the Expected Shortfall does.

1.5 Definition (Expected Shortfall).

Let Y be a random variable with corresponding distribution function F_Y satisfying $\mathbb{E}[(-Y) \vee 0] < \infty$.

The *lower tail Expected Shortfall of Y at level* $\alpha \in (0,1]$ is defined by

$$es_\alpha(Y) = es_\alpha(F_Y) = -\frac{1}{\alpha}\int_0^\alpha q^-_u(F_Y)\,\mathrm{d}u.$$

For $\alpha = 0$ define $es_0(F_Y) = -\operatorname{ess\,inf}(Y)$, where $\operatorname{ess\,inf}(Y)$ is the *essential infimum* given as $\operatorname{ess\,inf}(Y) = \sup\{C \in \mathbb{R} \mid \mathrm{P}(Y < C) = 0\}$.

As q^-_u and q^+_u can only differ for countably many points u, we can also write

$$es_\alpha(F_Y) = -\frac{1}{\alpha}\int_0^\alpha q^+_u(F_Y)\,\mathrm{d}u = \frac{1}{\alpha}\int_0^\alpha VaR_u(F_Y)\,\mathrm{d}u,$$

which demonstrates the connection to the Value at Risk. As the Value at Risk is monotonically decreasing in α by Remark 1.3, ii), the former representation shows $es_\alpha(Y) \geq VaR_\alpha(Y)$, so that the Expected Shortfall is more conservative. In addition, we can see here that small variations in α do not change es_α much – the map $\alpha \mapsto es_\alpha$ is continuous.

A further representation for the Expected Shortfall, for example shown in Acerbi and Tasche (2002b), is given by

$$\begin{aligned} es_\alpha &= -\alpha^{-1}\Big[\mathbb{E}\left[Y\,\mathbb{1}\left(Y \leq q^-_\alpha(F_Y)\right)\right] + q^-_\alpha(F_Y)\big(\alpha - \mathrm{P}(Y \leq q^-_\alpha(F_Y))\big)\Big] \\ &= -\alpha^{-1}\Big[\mathbb{E}\left[Y\,\mathbb{1}\left(Y \leq s\right)\right] + s\big(\alpha - \mathrm{P}(Y \leq s)\big)\Big] \qquad (1.2) \end{aligned}$$

for any $s \in [q^-_\alpha(F_y), q^+_\alpha(F_Y)]$, which explains the synonym *Tail Mean* used in the literature. Yet another term used for the Expected Shortfall is *Conditional Value at Risk* (see Corollary 4.3, Acerbi and Tasche (2002b), and Theorem 10, Rockafellar and Uryasev (2002)). Unfortunately this name suggests a representation of the Expected Shortfall as a conditional expectation, which in general does *not* exist. A more precise formulation of this fact is Corollary 6.2, Acerbi and Tasche (2002b).

The term $\alpha^{-1}\mathbb{E}\left[Y\,\mathbb{1}\left(Y \leq q_\alpha^-(F_Y)\right)\right]$ used in (1.2) is – quite intuitively – called *(lower) Tail Conditional Expectation*, with *lower* referring to the occurring lower quantile. The Tail Conditional Expectation is often also used synonymously with Expected Shortfall, raising some confusion, as the Tail Conditional Expectation is not coherent in general whereas the Expected Shortfall is advertised as such. Corollary 5.3, Acerbi and Tasche (2002b), gives precise conditions under which these risk measures agree; a proof for the coherence of the Expected Shortfall can be found in that reference as well.

Last for this section, we introduce risk measures obtained by adding several weighted Expected Shortfalls. Therefore let m be a probability measure on $[0,1]$, called *spectral measure*, and let $\kappa_m : \mathcal{L}_1 \to \mathbb{R}$,

$$\kappa_m(Y) = \int_0^1 es_\alpha(F_Y)\,\mathrm{d}m(\alpha) \tag{1.3}$$

be the *spectral risk measure associated to m*. Note that using $m = \delta_\alpha$ for some $\alpha \in (0,1]$ gives $\kappa_m = es_\alpha$. In general, the risk measure κ_m is law invariant as the Expected Shortfall has this property.

Calling such κ_m "spectral" is justified by a result of Kusuoka (2001), where it is shown that the law invariant, coherent and comonotonic risk measures are exactly the spectral risk measures associated to some probability measure m on $[0,1]$ defined in (1.3). Later, we consider the class of spectral risk measures with spectral measure having finite support which then is a finite convex combination of Expected Shortfalls for different levels.

The goals of the thesis regarding the risk measures so far are twofold. First, in Chapter 3 we deal with the asymptotic behaviour of the M-estimates $\widehat{q}_{n,\alpha}$ and $\widehat{es}_{n,\alpha}$ of q_α and es_α, respectively, if the size n of the sample grows (for details on M-estimates see Definition 1.13). Especially, we are interested in the weak convergence of the pair $\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha),\, b_n(\widehat{es}_{n,\alpha} - es_\alpha)\big)$ for properly chosen sequences $a_n, b_n \to \infty$. It is known (Knight, 2002) that the M-estimate for the quantile can have a convergence rate a_n slower than $\sqrt{n}$, dependent on the regularity of the distribution function in the considered quantile q_α. We shall answer the question whether the rate of convergence b_n for the Expected Shortfall estimator changes as well in the situation of low regularity around q_α. The M-estimator for the spectral risk measure κ_m is examined regarding the same question.

The second aim, achieved in Chapter 5, is to obtain the weak limit of the empirical quantile process $\alpha \mapsto \sqrt{n}\big(\widehat{q}_{n,\alpha} - q_\alpha\big)$ on a closed sub-interval of $(0,1)$ in an appropriate semimetric space. The theory of weak convergence of stochastic processes needed for this is detailed in Section 2.2.

1.4 Scoring functions

Here we introduce the concept of consistent scores, which is another way of identifying certain parameters. We shall state appropriate consistent scoring functions for the risk measures defined so far and prove a result concerning the existence of consistent scores for Bayes risks. Furthermore, we present the expectile risk measure which comprises benefits of the Value at Risk and the Expected Shortfall.

1.6 Definition (Scores).

Let $\mathcal{F}$ be a family of distribution functions.
A function $S : \mathbb{R}^k \times \mathbb{R} \to \mathbb{R}$, such that $S(x; \cdot)$ is $\mathcal{F}$-integrable is called *scoring function* or *score*; the function $S(x; F) = \mathbb{E}_F[S(x; Y)]$ is called *expected score.*
The score S is termed *$\mathcal{F}$-consistent* for the parameter $T : \mathcal{F} \to \mathbb{R}^k$ if $S(T(F); F) \leq S(x; F)$ for any $F \in \mathcal{F}$ and $x \in \mathbb{R}$. It is said to be *strictly $\mathcal{F}$-consistent* for T if it is $\mathcal{F}$-consistent and the equality $S(T(F); F) = S(x; F)$ implies $x = T(F)$ for every $F \in \mathcal{F}$ and $x \in \mathbb{R}$.
The parameter $T : \mathcal{F} \to \mathbb{R}^k$ is called *k-elicitable relative to $\mathcal{F}$* if there exists a strictly $\mathcal{F}$-consistent scoring function for T.

Using Bayesian terminology, the expected score is the Bayes risk, meaning that a functional T is elicitable relative to $\mathcal{F}$ if there is a strictly $\mathcal{F}$-consistent score, such that T is the Bayes predictor.

The most prominent questions to ask are whether there is any rich class $\mathcal{F}$, such that a parameter T is elicitable with respect to that, and how the strictly $\mathcal{F}$-consistent scores can be characterized. For example the mean is 1-elicitable with respect to the class of all distribution functions having finite first moment, where the class of consistent scoring functions is given by the Bregman-functions (Savage, 1971; Gneiting, 2011a). This can be generalized to vectors of rations of expectations (Fissler and Ziegel, 2016).

Quantiles are also understood well. Their score is important for us in the later chapters.

1.7 Example (Score for the quantile).
Let $\alpha \in (0, 1)$ be fixed and set $\mathcal{F}$ as the class of distribution functions with unique α-quantile. Then the α-quantile is elicitable. An $\mathcal{F}$-consistent scoring function is given by

$$S(x; z) = \big(\mathbb{1}\,(z \leq x) - \alpha\big)\big(g(x) - g(z)\big) \tag{1.4}$$

for an increasing function g. If g is strictly increasing, then S in (1.4) is strictly $\mathcal{F}$-consistent.

Conversely, choosing any strictly consistent score S for the α-quantile, it is necessarily of the form above under mild regularity conditions (Gneiting, 2011b).

The most prominent example arises when choosing $g(x) = x$ which yields a piecewise linear function with slope $-\alpha$ or $1 - \alpha$ depending on which side of y we are. This is the so called *check function*, also referred to as *tick, pinball* or *hinge loss.* ◇

Steinwart et al. (2014) gives equivalent characterizations of elicitability for *continuous*

1-dimensional parameters. An important characterization uses the level sets of the functional. We shall call the level sets of a property T *convex* if whenever $F_0, F_1 \in \mathcal{F}$ and $p \in (0,1)$ such that $F_p = (1-p)F_0 + pF_1 \in \mathcal{F}$ holds, $t \in T(F_0)$ together with $t \in T(F_1)$ implies $t \in T(F_p)$. Corollary 9, Steinwart et al. (2014), states that convexity of the level sets is equivalent to elicitability of T.

Dropping the continuity assumption, the question of a general characterization for elicitability remains open. On the one hand, Osband (1985) shows that convexity of the level sets remains a necessary condition for T being elicitable; see also Theorem 6, Gneiting (2011a). On the other hand, the mode functional is not elicitable, despite having convex level sets (Heinrich, 2014), hence, the condition of convex level sets is not sufficient for elicitability.

Showing non-convexity of level sets is a famous way to show non-elicitability of parameters. For example, Theorem 11, Gneiting (2011a), shows that the Expected Shortfall does not have convex level sets provided $\mathcal{F}$ contains at least all measures with finite support or the mixtures of the absolutely continuous distributions with compact support. Thus, the Expected Shortfall cannot be elicitable. The variance can be treated in the same way. These negative results do not purport that the respective parameter is not elicitable with respect to *any* class $\mathcal{F}$. Rather, what should be understood when talking about non-elicitability is that the parameter is not elicitable with respect to an "acceptably rich" class of distribution functions.

Additionally, it is sometimes benefiting to change from 1-elicitability to k-elicitability instead, as some non-elicitable parameters are one entry of a k-dimensional parameter, which is elicitable. For example, the variance is jointly elicitable with the mean, and likewise the Expected Shortfall is 2-elicitable when simultaneously considering the Value at Risk. More general this holds for spectral risk measures as in the following example.

1.8 Example (k-elicitability of spectral risk measures associated to m). Fissler and Ziegel (2016) just showed that the Expected Shortfall is jointly elicitable with the Value at Risk. More general, spectral risk measures κ_m with spectral measure m having finite support on $(0,1]$ are s-elicitable for $s=1$ or $s=k+1$ with respect to the class of distribution functions having finite first moment and unique quantiles (Fissler and Ziegel, 2016, Corollary 5.4). In the latter case, writing $m = \sum_{l=1}^k p_l \delta_{\alpha_l}$ for $\alpha_1, \ldots, \alpha_k \in (0,1]$ and $p_l \in (0,1)$ with $\sum_{l=1}^k p_l = 1$, a strictly consistent scoring function for the vector $(q_{\alpha_1}, \ldots, q_{\alpha_k}, \kappa_m)$ is given by

$$S(x_1, \ldots, x_k, x_{k+1}; z) = \sum_{l=1}^{k} \Bigg(\Big(1 + \frac{p_l}{\alpha_l} G(x_{k+1})\Big) \left(\mathbb{1}\left(z \le x_l\right) - \alpha\right) \left(x_l - z\right) \\ + p_l\big(G(x_{k+1})(x_{k+1} - z) - \mathcal{G}(x_{k+1})\big) \Bigg),$$

where $\mathcal{G}$ is a three-times continuously differentiable function, $G = \mathcal{G}'$, and it is required that $G' > 0$. From the proof of Corollary 5.5 in Fissler and Ziegel (2016), we may choose $\mathcal{G}$ so that $\lim_{x\to-\infty} G(x) = 0$. For the pair (q_α, es_α), $\alpha \in (0,1)$, the above score reduces to

$$\begin{aligned} S(x_1, x_2; z) = \big(\mathbb{1}(z \le x_1) - \alpha\big)(x_1 - z) + G(x_2)\left(x_2 + \alpha^{-1}\big(\mathbb{1}(z \le x_1) - \alpha\big)(x_1 - z)\right) \\ - \mathcal{G}(x_2) - G(x_2) z. \end{aligned} \tag{1.5}$$

The case $s = 1$ can be chosen if $m = \delta_1$, as then $\kappa_m = \mathbb{E}[Y]$ is valid and the mean is 1-elicitable. ◇

The score in (1.5) is one of the main tools in Chapter 3 where it is used to define the M-estimators for the quantile and Expected Shortfall as well as deduce their limiting behaviour.

We observe that, when knowing the α-quantile, we can set $x_1 = q_\alpha$ in (1.5), such that the first part of S vanishes after taking the expectation, whereas the second part becomes a "score" for the Expected Shortfall, namely

$$es_\alpha = es_\alpha(F) = \operatorname*{arg\,min}_{x_2 \in \mathbb{R}} \mathbb{E}_F\big[S(q_\alpha, x_2; Y)\big]. \tag{1.6}$$

This is the *conditional* elicitability of the Expected Shortfall as introduced in Emmer et al. (2015). Let us write

$$S_0(x_1; z) = \alpha^{-1}\big(\mathbb{1}(z \le x_1) - \alpha\big)(x_1 - z)$$

and observe that S_0 is a score for the quantile q_α. Using Corollary 5.4, Fissler and Ziegel (2016), the function

$$S(x_1, x_2; z) = S_0(x_1; z) + G(x_2)(x_2 + S_0(x_1; z)) - (\mathcal{G}(x_2) - \mathcal{G}(z))$$

elicits (q_α, es_α) provided $\mathcal{G}$ is an $\mathcal{F}$-integrable, three-times continuously differentiable function, $\mathcal{G}' = G$ holds, and $G' > 0$. Assuming we know the α-quantile, we can insert it for x_1 above such that after taking expectations with respect to F we end up with

$$S(q_\alpha, x_2; F) = S_0(q_\alpha; F) + G(x_2)(x_2 + S_0(q_\alpha; F)) - \big(\mathcal{G}(x_2) - \mathbb{E}_F[\mathcal{G}(Y)]\big).$$

Quite conveniently this gives a differentiable function which can be minimized by setting the derivative equal to 0. Doing this we have to solve $0 = G'(x_2)\big(x_2 + S_0(q_\alpha; F)\big)$ which is fulfilled by $x_2^* = -S_0(q_\alpha; F)$ as $G' > 0$. By (1.6) the minimizer x_2^* has to equal es_α, hence $es_\alpha = -S_0(q_\alpha; F)$. This shows that the Expected Shortfall can be seen as the Bayes risk of S_0.

The scheme prescribed here can also be applied to more general Bayes risks, so that when starting with a parameter T, which is k-elicitable with respect to some class $\mathcal{F}$ with consistent score S_0, the pair $(T, S_0(T; F))$ is $(k+1)$-elicitable. This provides our first result which also clarifies Frongillo and Kash (2015a, Corollary 1); the proof is given in Section 1.6..

1.9 Theorem.

Let $S_0(x;z)$ and $S_1(x;z)$ be $\mathcal{F}$-consistent scoring functions for $T : \mathcal{F} \to \mathbb{R}^k$ and set $\gamma(F) = -\min_{x\in\mathbb{R}^k} S_0(x;F)$, provided the minimum is attained. Assume that G is strictly increasing. Further define the function $\mathcal{G}$ such that $\mathcal{G}' = G$ and assume that $\mathcal{G}$ is $\mathcal{F}$-integrable. If S_0 or S_1 is strictly $\mathcal{F}$-consistent, then

$$S(x_1, x_2; z) = S_1(x_1; z) + G(x_2)(x_2 + S_0(x_1; z)) - (\mathcal{G}(x_2) - \mathcal{G}(z))$$

is strictly $\mathcal{F}$-consistent for (T, γ).

Let us now turn to the next risk measure dealt with later on. We saw that the quantile is the Bayes rule under the asymmetric piecewise linear score in Example 1.7. When considering an asymmetric piecewise *quadratic* score for distribution functions with finite second moment, $S(x;z) = |\mathbb{1}(x \geq z) - \tau|(x-z)^2$, $\tau \in (0,1)$, the resulting Bayes rule is called *expectile*. This asymmetric score was first proposed by Newey and Powell (1987) to design tests for homoscedasticity and symmetry of the error distribution in a linear regression setting, their idea coming from analogous tests constructed using quantiles. We use a slightly different scoring function which only needs a finite first moment of the considered distribution function.

1.10 Definition (Expectile).

Let $Y \in \mathcal{L}_1$ with distribution function F and $\tau \in (0,1)$.
The τ-*expectile* or *expectile at level* τ, denoted with $\mu_\tau(F)$, is the unique minimizer of $x \mapsto \mathbb{E}_F[S_\tau(x;Y)]$ where

$$S_\tau(x;z) = \frac{\tau}{2}\left[((z-x)^+)^2 - (z^+)^2\right] + \frac{1-\tau}{2}\left[((z-x)^-)^2 - (z^-)^2\right].$$

Another possibility to define expectiles is to say that $\mu_\tau = \mu_\tau(F)$ is the unique value fulfilling

$$\frac{\tau}{1-\tau} = \frac{\int_{(-\infty,x)} (x-y)\,\mathrm{d}F(y)}{\int_{(x,\infty)} (y-x)\,\mathrm{d}F(y)}, \quad x \in \mathbb{R}.$$

This shows the dependence of μ_τ on *both* tails of the distribution function. On the other hand that defining equation is similar to

$$\frac{\alpha}{1-\alpha} = \frac{F(x)}{1-F(x)}, \quad x \in \mathbb{R},$$

which is solved uniquely by the quantile $q_\alpha(F)$ and thus explains the name “expectile” as an acronym based on *expec*tation and quan*tile*.

The risk measure induced by expectiles as proposed by Kuan et al. (2009) is called *Expectile-based Value at Risk* and is defined by

$$EVaR_\tau(Y) = -\mu_\tau(F).$$

In the literature, the Expectile-based Value at Risk is deemed less common and more conservative than the Value at Risk (for example Jones (1994)). Until recently the major criticism of utilizing the expectile as a risk measure has been that only downward deviations should contribute to the risk. An economic justification for also taking upward deviations into account was just given in Ehm et al. (2016).

The name $EVaR$ hints at a similarity to the Value at Risk, already conjectured in Newey and Powell (1987) on page 824 as "expectiles have properties that are similar to quantiles". Jones (1994) subsequently showed that expectiles are in fact quantiles of a transformation of the underlying distribution function explaining the similar properties. This connection makes it possible to access quantiles and even the Expected Shortfall with help of expectiles, see Taylor (2008). However, while quantiles in general lack the property of coherence, expectiles were shown to be the only law invariant, elicitable and coherent risk measures (Bellini and Bignozzi, 2015, Theorem 4.9).

By definition expectiles are elicitable relative to the class of distribution functions having finite first moment; the class of consistent scoring functions is characterized in Theorem 10, Gneiting (2011a).

Regarding expectiles at a fixed level, we last want to mention that they are not comonotone additive (Bellini and Bignozzi, 2015) which is somewhat diminishing the delight of using expectiles.

In Chapter 5 we consider the empirical expectile process $\tau \mapsto \sqrt{n}\big(\mu_\tau(F_n) - \mu_\tau(F)\big)$ on a closed subinterval of $(0,1)$. As the map $\tau \mapsto \mu_\tau(F)$ is continuous for all distribution functions F with finite first moment (Holzmann and Klar, 2016), the former process attains values in the space of continuous functions almost surely. On the other hand, Holzmann and Klar (2016) also show that for $\tau_0 \in (0,1)$ the weak limit of the random variable $\sqrt{n}\big(\mu_{\tau_0}(F_n) - \mu_{\tau_0}(F)\big)$ is non-normal if F is discontinuous in μ_{τ_0}. Thus, the continuous empirical expectile process must converge to a discontinuous limit in such a situation. As a main result for the thesis we deduce the weak limit of the empirical expectile process in the sense of Section 2.2 under mild assumptions on F.

For the last part of this section we are interested in comparing possible estimates for an elicitable parameter T with consistent score S. Having chosen two estimators T^1 and T^2 for T, we retrospective want to know, which estimation procedure is "superior"? One possibility for answering this is to turn the attention to the potential economic loss which occurs when favouring one estimator above the other. As mentioned above, we can view S as an indicator for the economic loss and thus have to compare the values $S(T^1; F_n)$ and $S(T^2; F_n)$. We then prefer the estimator which yields the lower value.

There is always more than one consistent score for T, and choosing another one could rank T^1 and T^2 differently; see for example Patton (2016). This leads to the question whether $S(T^1; F_n)$ is smaller or bigger than $S(T^2; F_n)$ for every possible score S – or at least for a reasonably large class $\mathcal{S}$ thereof. Therefore it is convenient to parametrize the class $\mathcal{S}$. Assume that we can find a family $(S_v)_{v\in\mathbb{V}}$ of *elementary scores* such that for every score $S \in \mathcal{S}$ there exists a measure λ_S on $\mathbb{V}$ with

$$S(x; y) = \int_{\mathbb{V}} S_v(x; y) \,\mathrm{d}\lambda_S(v).$$

This representation is referred to as *decomposition of S in elementary scores S_v*. If now $S_v(T^1; F_n)$ and $S_v(T^2; F_n)$ are ordered alike for every $v \in \mathbb{V}$, this order will translate to $S(T^1; F_n)$ and $S(T^2; F_n)$. Hence, by comparing the values of the elementary scores, we can deduce the dominance of one estimation scheme over the other.

This approach is mostly used for assessing the accuracy of predictions for future values. The topic probably started with Diebold and Mariano (1995) and recently attracted attention; see for example Clark and McCracken (2001), Ehm et al. (2016), Ziegel et al. (2017) and references therein. As proven in Theorem 1, Ehm et al. (2016), a subset of the scores for the Value at Risk and the expectiles, respectively, are decomposable in the above sense. Our contribution to that field is the following result which in fact is a corollary from Theorem 1.9. The proofs of both following results are deferred to Section 1.6.

1.11 Theorem.

Let $S_0(x_1; z)$ and $S_1(x_1; z)$ be consistent scoring functions for the functional $T : \mathcal{F} \to \mathbb{R}^k$ and define the property $\gamma(F) = -\min_{x_1\in\mathbb{R}^k} S_0(x_1; F)$. Then there is a score S for (T, γ), which admits a partial decomposition in elementary scores, namely the function

$$S(x_1, x_2; z) = S_1(x_1; z) + \int S_{v_2}(x_1, x_2; z) \,\mathrm{d}\, G(v_2)$$

is a consistent scoring function for (T, γ). Here

$$S_{v_2}(x_1, x_2; z) = \mathbb{1}\,(v_2 \le x_2)\ (v_2 + S_0(x_1; z)) + \mathbb{1}\,(v_2 \le z)\ (z - v_2)$$

are the elementary scores and $\mathrm{d}\, G$ is a measure which is finite on all intervals of the form $(-\infty, x]$, $x \in \mathbb{R}$. The above score S is strictly consistent for (T, γ) if $\mathrm{d}\, G$ puts positive mass on all open intervals.

This theorem immediately yields a decomposition for consistent scores of risk measures comprised of a Bayes rule and the associated Bayes risk, which is a generalization of Proposition 2.1, Ziegel et al. (2017).

1.12 Corollary.

Let $S_0(x_1; z)$ be a consistent scoring function for the functional $T : \mathcal{F} \to \mathbb{R}^k$ and define the property $\gamma(F)$ as in Theorem 1.11. Assume that T has a scoring function $S_1(x_1; z)$ which admits a decomposition in elementary scores,

$$S_1(x_1; z) = \int S_{v_1}(x_1; z) \,\mathrm{d}\, H(v_1),$$

for a locally finite measure H. Then there is a consistent score S for (T, γ), which is decomposable in elementary scores. Especially, with S_{v_2} and $\mathrm{d}\, G$ as in Theorem 1.11 it holds that

$$S(x_1, x_2; z) = \int S_{v_1}(x_1; z) \,\mathrm{d}\, H(v_1) + \int S_{v_2}(x_1, x_2; z) \,\mathrm{d}\, G(v_2).$$

1.5 M-Estimation

In this section we briefly introduce M-estimators which are the main estimators used in the present work. After defining them we cite a consistency result which motivates a new result concerning the consistency of M-estimators where nuisance parameters are present.

A comprehensive collection of classical results on M-estimation can be found in van der Vaart (1998), Chapter 5. In order to better relate the discussion and new results to the existing theory we change the notation in this chapter, writing $m_x(z) = S(x; z)$ for scores and $\vartheta_0 = T$ for the parameter of interest.

In the former section we identified an elicitable parameter ϑ_0 of a distribution function as minimizers of some deterministic function. The latter was obtained by integrating the consistent score m_x with respect to the distribution function F. Having an (independent identically distributed) sample $Y_1, \ldots, Y_n$ drawn from F, we can use the empirical distribution function F_n as an estimate for F and consider $m_x(F_n) = S(x; F_n)$. Under integrability conditions on m_x, $m_x(F_n)$ is a strongly consistent estimator for $m_x(F)$ by the strong law of large numbers and we hope that this translates to the respective sequence of minimizers. Precisely, we await the minimizers of the map $x \mapsto m_x(F_n)$ to (exist and) be useful estimators for $\vartheta_0 = \arg\min_x m_x(F)$.

1.13 Definition.

An estimator ϑ_n is called *M-estimator* if it fulfils

$$\vartheta_n \in \arg\min_x \; m_x(F_n) = \arg\min_x \frac{1}{n} \sum_{i=1}^{n} m_x(Y_i).$$

The function $m_x(F_n)$ is mostly called *empirical criterion function* or *empirical contrast* in the context of M-estimation; $m_x(F)$ is named *asymptotic criterion function* or *asymptotic contrast.*

In the following part we often use the *outer* expectation $\mathbb{E}^\circ$ and the *outer* probability P°, which help to overcome measurability issues. The definitions thereof – together with the theory of weak convergence for random processes needed at the end of this section – is shifted to Chapter 2.

The idea in M-estimation is to deduce the behaviour of the implicitly defined estimator ϑ_n by using assertions about the behaviour of $m_x(F_n)$ as estimator for $m_x(F)$. We actually expect that if $m_x(F_n)$ is close to $m_x(F)$, the distance of their minimizers is small in probability. To give an idea, we state one of the best known consistency results in the present context. In Theorem 5.7, van der Vaart (1998), it is proven that if

$$\sup_{\vartheta} |m_\vartheta(F_n) - m_\vartheta(F)| = o_{\mathrm{P}}(1),$$

$$\sup_{d(\vartheta,\vartheta_0)\geq\varepsilon} m_\vartheta(F) > m_{\vartheta_0}(F)$$

holds for any $\varepsilon > 0$, then any sequence of M-estimators ϑ_n converges to ϑ_0. Here the supremum is taken over $\vartheta \in \Theta$ for a metric space (Θ, d) with $\vartheta_0 \in \Theta$.

The first assumption says that the empirical contrast has to converge uniformly in probability to the asymptotic one. Heuristically, the estimate $m_x(F_n)$ then is similarly close to $m_x(F)$ for any point x. The second statement means that the true minimizer has to be *well-separated*: Whenever we determine the value of the asymptotic contrast outside a small neighbourhood of the true minimizer, we only obtain bigger values of the contrast.

The assertion before only takes care of consistency of the M-estimator ϑ_n. In order to deduce the rate of convergence we need a different pair of assumptions, namely

$$\inf_{d(\vartheta,\vartheta_0)<\delta} \mathbb{E}\left[m_{\vartheta_0}(Y) - m_\vartheta(Y)\right] \geq C\delta^\alpha$$

$$\mathbb{E}^\circ\left[\sup_{d(\vartheta,\vartheta_0)<\delta} \left|\mathbb{G}_n[m_\vartheta(Y) - m_{\vartheta_0}(Y)]\right|\right] \leq C\delta^\beta$$

for every sufficiently small $\delta > 0$, a constant $C > 0$ and $\alpha > \beta$. If additionally the sequence ϑ_n of M-estimators is consistent for ϑ_0, then $n^{1/(2\alpha-2\beta)}\, d(\vartheta_n, \vartheta_0)$ is bounded in outer probability (van der Vaart, 1998, Theorem 5.52). Thus, the rate of convergence is $n^{1/(2\alpha-2\beta)}$ if α and β were chosen optimally.

Assume now that ϑ_n and ϑ_0 are some entries of a k-dimensional vector $e_n = (\vartheta_n, \eta_n)$ and $e_0 = (\vartheta_0, \eta_0)$, respectively, with $k > 1$. Then usage of the former assertion forces every entry of e_n to have the same rate of convergence. We overcome this limitation with the next theorem by considering some entries of e_n as nuisance parameters. This is similar to van der Vaart (1998, Theorem 5.52).

1.14 Theorem.

Assume that $(\Theta_0, d_0), (\Theta_1, d_1)$ *are metric spaces and that for all* $\eta \in \Theta_0$, $\vartheta \in \Theta_1$, *the map* $y \mapsto m_{\eta,\vartheta}(y)$ *is measurable. Further, suppose that for fixed* $C > 0$ *and* $\alpha > \beta$, *every* $n \in \mathbb{N}$ *and all sufficiently small* $\varepsilon, \delta > 0$ *it holds that*

$$\inf_{d_0(\eta,\eta_0)\le\varepsilon} \inf_{d_1(\vartheta,\vartheta_0)\ge\delta} \mathbb{E}\left[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\right] \ge C\delta^\alpha \tag{1.7}$$

and

$$\mathbb{E}^\circ\left[\sup_{d_0(\eta,\eta_0)\le\varepsilon} \sup_{d_1(\vartheta,\vartheta_0)\le\delta} \left|\mathbb{G}_n\left[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\right]\right|\right] \le C\delta^\beta. \tag{1.8}$$

Additionally, presume that η_n *converges to* η_0 *in outer probability and* ϑ_n *converges to* ϑ_0 *in outer probability and fulfils*

$$\mathbb{E}_n\Big[m_{\eta_n,\vartheta_n}(Y)\Big] \le \mathbb{E}_n\Big[m_{\eta_n,\vartheta_0}(Y)\Big] + O_{\mathrm{P}}\big(n^{-\alpha/(2(\beta-\alpha))}\big).$$

Then $n^{1/(2\alpha-2\beta)}d_1(\vartheta_n, \vartheta_0)$ *is bounded in outer probability.*

The last condition in the former theorem says that the considered sequence ϑ_n of estimators only needs to be an approximate minimizer of the empirical contrast. The proof of the theorem is given in Section 1.6.

After having obtained the rate of convergence a_n of an estimator ϑ_n, we are interested in the weak limit of the sequence $a_n(\vartheta_n - \vartheta_0)$. In the context of M-estimators this can be done via the argmax-continuity theorem where the weak convergence of the properly rescaled and translated empirical contrast process is transferred to the weak convergence of the minimizers (van der Vaart, 1998, Theorem 5.56 or Corollary 5.58).

In Chapters 3 and 4 we consider situations of M-estimators ϑ_n for which the empirical contrast functions do *not* converge in distribution. The problem arises as the ϑ_n are one entry of k-dimensional vectors e_n, $k > 1$, where all entries of e_n can have different rates of convergence. We deal with this more generally using the next assertion. This at first compares the asymptotic behaviour of M-estimators; however, it enables us to separate the entries of e_n, such that working with the argmax-continuity theorem under different rates of convergence becomes possible again.

1.15 Lemma.

Let M_n *and* M_n' *be real valued processes where the process* M_n' *admits the representation* $M_n'(u) = N_n(u) + R_n$, $u \in \mathbb{R}^k$. *Here,* R_n *is a sequence of random variables not* *depending on u. Assume that*

$$\sup_{u\in K}\big|M_n(u)-M_n'(u)\big| = o_{\mathrm{P}}(1) \tag{1.9}$$

and that $N_n \rightsquigarrow N$ holds in $\big(\ell^\infty(K), \|\cdot\|_K\big)$ for every compact set $K\subset\mathbb{R}^k$ and some process N. Choose $(\vartheta_n, u_n)\in\mathbb{R}^{2k}$ as minimizer of (M_n, M_n') and in addition assume that $\vartheta_n = O_{\mathrm{P}}(1)$ and $u_n \xrightarrow{\mathcal{L}} u_0$ as variables in $\mathbb{R}^k$ where u_0 is the unique minimizer of N (assuming all of these variables exist). Then it holds that $\vartheta_n = u_n + o_{\mathrm{P}}(1)$.

Note that the approximating processes M_n' converge apart from a sequence of random variables R_n where the latter is not important for determining the minimizers u_n. We prove the lemma in Section 1.6.

1.6 Proofs

This section contains the proofs for the statements in the present chapter.

1.6.1 Proofs for Section 1.4

We start by proving the assertions considering the existence of consistent scores for Bayes risks and their decomposition in elementary scores as stated in Theorems 1.9 and 1.11 as well as Corollary 1.12.

Proof of Theorem 1.9. We show both parts simultaneously; the reasoning is analogous to Frongillo and Kash (2015a, Corollary 1). By the subgradient inequality for strictly increasing G any $x_2 \neq -S_0(x_1;F)$ fulfils

$$\begin{aligned}
&S\big(x_1, -S_0(x_1;F);F\big)\\
&\quad = S_1(x_1;F) - \big(\mathcal{G}(-S_0(x_1;F)) - \mathcal{G}(F)\big)\\
&\quad < S_1(x_1;F) + G(x_2)\big(x_2 + S_0(x_1;F)\big) - \big(\mathcal{G}(x_2) - \mathcal{G}(F)\big) = S\big(x_1, x_2;F\big),
\end{aligned}$$

so that the minimum of $x_2 \mapsto S\big(x_1,x_2;F\big)$ is uniquely determined by $x_2^* = -S_0(x_1;F)$. Letting $\widetilde{S}(x_1;F) = S\big(x_1, x_2^*;F\big)$, we have

$$\arg\min_{x_1}\widetilde{S}(x_1;F) = \arg\min_{x_1}\Big(S_1(x_1;F) - \mathcal{G}(-S_0(x_1;F))\Big)$$

as a first step. But since S_0 and S_1 have at least one common minimizer, it follows that

$$\arg\min_{x_1}\widetilde{S}(x_1;F) = \arg\min_{x_1}\Big(S_1(x_1;F)\Big) \cap \arg\min_{x_1}\Big(-\mathcal{G}(-S_0(x_1;F))\Big).$$

As $\mathcal{G}$ is strictly increasing, this reduces to

$$\begin{aligned}\arg\min_{x_1} \widetilde{S}(x_1;F) &= \arg\min_{x_1} \Big(S_1(x_1;F)\Big) \cap \arg\max_{x_1} \Big(-S_0(x_1;F)\Big)\\ &= \arg\min_{x_1} \Big(S_1(x_1;F)\Big) \cap \arg\min_{x_1} \Big(S_0(x_1;F)\Big).\end{aligned}$$

Thus, it follows that

$$\arg\min_{x_1,x_2} S(x_1,x_2;F) = \Big\{(t,\, -S_0(t;F)) \;\mid\; t \in \arg\min_{x_1} \left(S_1(x_1;F)\right) \cap \arg\min_{x_1} \left(S_0(x_1;F)\right)\Big\}.$$

The set on the right hand side reduces to a unique pair if $|\arg\min_{x_1} \left(S_1(x_1;F)\right)| = 1$ or $|\arg\min_{x_1} \left(S_0(x_1;F)\right)| = 1$, which is the assertion of the theorem. □

The next proof is inspired by the one of Proposition 2.1, Ziegel et al. (2017).

Proof of Theorem 1.11. By Theorem 1.9, the pair (T,γ) has consistent score

$$S(x_1,x_2;z) = S_1(x_1;z) + G(x_2)(x_2 + S_0(x_1;z)) - (\mathcal{G}(x_2) - \mathcal{G}(z)) \tag{1.10}$$

for G and $\mathcal{G}$ given in Theorem 1.9. Especially, the function G is increasing and positive, and $\mathcal{G}$ is differentiable with $\mathcal{G}' = G$. Observe that G can be written as

$$G(x_1) = \int \mathbb{1}\left(v \le x_1\right)\, \mathrm{d}\, G(v)$$

for the induced locally finite measure $\mathrm{d}\, G$. Next integrate the proposed S_{v_2} with respect to $\mathrm{d}\, G$ to obtain

$$\begin{aligned}\int S_{v_2}(x_1,x_2;z)\,\mathrm{d}\, G(v_2) &= S_0(x_1;z)\int \mathbb{1}\left(v_2 \le x_2\right)\,\mathrm{d}\, G(v_2) + z\int \mathbb{1}\left(v_2 \le z\right)\,\mathrm{d}\, G(v_2)\\ &\quad + \int v_2\left(\mathbb{1}\left(v_2 \le x_2\right) - \mathbb{1}\left(v_2 \le z\right)\right)\mathrm{d}\, G(v_2)\\ &= S_0(x_1;z)\, G(x_2) + z\, G(z) + x_2\, G(x_2) - z\, G(z) - \int_z^{x_2} G(v_2)\,\mathrm{d}v_2\\ &= G(x_2)\left(x_2 + S_0(x_1;z)\right) - \left(\mathcal{G}(x_2) - \mathcal{G}(z)\right)\end{aligned}$$

with a partial integration for the second equality. We insert this in (1.10) to obtain the stated representation of S.

Next note that, by the positivity of G, the measure $\mathrm{d}\, G$ puts finite mass on every interval of the form $(-\infty, x]$. Last, the score in (1.10) is strictly consistent if G is strictly increasing (see Theorem 1.9). In that case the measure $\mathrm{d}\, G$ puts positive mass on any open interval in $\mathbb{R}$. □

1.6.2 Proofs for Section 1.5

We start this section with proving Theorem 1.14 which generalizes the proof of van der Vaart (1998, Theorem 5.52).

Proof of Theorem 1.14. We set $r_n = n^{1/(2\alpha-2\beta)}$ and suppose that ϑ_n minimises the map $\vartheta \mapsto \mathbb{E}_n\big[m_{\eta_n,\vartheta}(Y)\big]$ up to a random variable $R_n = O_{\mathrm{P}}(r_n^{-\alpha})$.

For each n the set $\Theta_1 \setminus \{\vartheta_0\}$ can be partitioned into the sets

$$D_{j,n} = \{\vartheta \,|\, 2^{j-1} < r_n d_1(\vartheta, \vartheta_0) \le 2^j\}, \quad j \in \mathbb{Z}.$$

If $r_n d_1(\vartheta_n, \vartheta_0) > 2^L$ for some $L \in \mathbb{Z}$, then ϑ_n must be in one of the $D_{j,n}$ for $j \ge L$. Further, if $\gamma > 0$ and $d_1(\vartheta_n, \vartheta_0) \le \frac{\gamma}{2}$, then $\vartheta_n \in D_{j,n}$ for $2^j \le \gamma\, r_n$. This gives

$$\begin{aligned}\mathrm{P}^\circ\Big(r_n d_1(\vartheta_n, \vartheta_0) > 2^L\Big) &\le \mathrm{P}^\circ\Bigg(\bigcup_{\substack{j\ge L\\ 2^j \le \gamma\, r_n}} \{\vartheta_n \in D_{j,n}\} \cap \{d_1(\vartheta_n, \vartheta_0) \le \tfrac{\gamma}{2}\} \cap \{d_0(\eta_n, \eta_0) \le \gamma\}\Bigg)\\ &\quad + \mathrm{P}^\circ\big(2d_1(\vartheta_n, \vartheta_0) > \gamma\big) + \mathrm{P}^\circ\big(d_0(\eta_n, \eta_0) > \gamma\big).\end{aligned}$$

Assume $\vartheta_n \in D_{j,n}$ for a j involved in the above union. Then, by assumption on ϑ_n, the infimum of the map $\vartheta \mapsto \mathbb{E}_n\big[m_{\eta_n,\vartheta}(Y) - m_{\eta_n,\vartheta_0}(Y)\big]$ over $D_{j,n}$ is at most R_n. If we suppose that in addition $d_0(\eta_n, \eta_0) \le \gamma$ holds, then the infimum of the map $(\eta, \vartheta) \mapsto \mathbb{E}_n\big[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\big]$ over $\mathrm{cl}(B_\gamma(\eta_0)) \times D_{j,n}$ is smaller than R_n as well. Hence, if $r_n^\alpha R_n \le C'$ for some $C' < \infty$, this infimum is smaller than $\frac{C'}{r_n^\alpha}$. Thus,

$$\begin{aligned}\mathrm{P}^\circ\Big(r_n d_1(\vartheta, \vartheta_0) > 2^L\Big) &\le \sum_{\substack{j\ge L\\ 2^j \le \gamma\, r_n}} \mathrm{P}^\circ\Big(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))} \inf_{\vartheta\in D_{j,n}} \mathbb{E}_n\big[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\big] \le \tfrac{C'}{r_n^\alpha}\Big)\\ &\quad + \mathrm{P}^\circ\big(r_n^\alpha R_n > C'\big) + \mathrm{P}^\circ\big(2d_1(\vartheta_n, \vartheta_0) > \gamma\big) + \mathrm{P}^\circ\big(d_0(\eta_n, \eta_0) > \gamma\big)\end{aligned} \tag{1.11}$$

follows. Observe that the last three summands can be made small for any $\gamma > 0$ by choosing n and C' big enough, as $R_n = O_{\mathrm{P}}(r_n^{-\alpha})$ and $\eta_n \to \eta$ as well as $\vartheta_n \to \vartheta_0$ in outer probability by assumption.

Now choose $\gamma > 0$ small enough to ensure that the conditions of the theorem hold for all $\delta, \varepsilon \le \gamma$. Every j involved in the above sum fulfils $\frac{2^j}{r_n} \le \gamma$, so that assumption (1.7) leads to

$$\begin{aligned}&\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))} \inf_{\vartheta\in D_{j,n}} \mathbb{E}\big[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\big]\\ &\ge \inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))} \inf_{\gamma \ge d_1(\vartheta,\vartheta_0) \ge \frac{2^{j-1}}{r_n}} \mathbb{E}\big[m_{\eta,\vartheta}(Y) - m_{\eta,\vartheta_0}(Y)\big] \ge C\left(\frac{2^{j-1}}{r_n}\right)^\alpha.\end{aligned} \tag{1.12}$$

Hence,

$$
\begin{aligned}
&\sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Big(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}} \mathbb{E}_n\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\leq \tfrac{C'}{r_n^\alpha}\Big)\\
&= \sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Bigg(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}} \Big((\mathbb{E}_n-\mathbb{E})\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\\
&\qquad\qquad +\mathbb{E}\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\Big)\leq \tfrac{C'}{r_n^\alpha}\Bigg)\\
&\leq \sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Bigg(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}} \Big((\mathbb{E}_n-\mathbb{E})\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\Big)\\
&\qquad\qquad +\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}}\mathbb{E}\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\leq \tfrac{C'}{r_n^\alpha}\Bigg)\\
&\leq \sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Bigg(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}} (\mathbb{E}_n-\mathbb{E})\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\leq \frac{\big(C'-C\,2^{(j-1)\alpha}\big)}{r_n^\alpha}\Bigg)
\end{aligned}
$$

is valid, where the first inequality needs Lemma 2.24 below and the last inequality uses (1.12). We now choose L large enough to guarantee $C'\leq C\,2^{(L-1)\alpha-1}$ so that

$$
C'-C\,2^{(j-1)\alpha}\leq C\,2^{(j-1)\alpha-1}-C\,2^{(j-1)\alpha}=-C\,2^{(j-1)\alpha-1}
$$

holds for $j\geq L$. This means that the former sum does not exceed

$$
\sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Big(\inf_{\eta\in\mathrm{cl}(B_\gamma(\eta_0))}\inf_{\vartheta\in D_{j,n}} (\mathbb{E}_n-\mathbb{E})\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big)\leq \tfrac{-C\,2^{(j-1)\alpha}}{2\,r_n^\alpha}\Big).
$$

By taking absolute values and multiplying with $\sqrt{n}$ this expression is smaller than (or equal to)

$$
\sum_{\substack{j\geq L\\ 2^j\leq \gamma\, r_n}} \mathrm{P}^\circ\Bigg(\sup_{(\eta,\vartheta)\in B}\Big|\mathbb{G}_n\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\Big|\geq C\sqrt{n}\,\frac{2^{(j-1)\alpha}}{r_n^\alpha}\Bigg)
$$

with $B=\mathrm{cl}(B_\gamma(\eta_0))\times\mathrm{cl}(B_{2^j/r_n}(\vartheta_0))$, where we also used $D_{j,n}\subset\mathrm{cl}\big(B_{2^j/r_n}(\vartheta_0)\big)$. Due to

Markov's inequality and the assumption (1.8) this term is finally not bigger than

$$\sum_{\substack{j\geq L\\ 2^j\leq\gamma r_n}} \frac{r_n^\alpha}{C\sqrt{n}\,2^{(j-1)\alpha}}\mathbb{E}^\circ\Big[\sup_{(\eta,\vartheta)\in B}\big|\mathbb{G}_n\big[m_{\eta,\vartheta}(Y)-m_{\eta,\vartheta_0}(Y)\big]\big|\Big] \leq \sum_{\substack{j\geq L\\ 2^j\leq\gamma r_n}} \frac{\big(\frac{2^j}{r_n}\big)^\beta r_n^\alpha}{\sqrt{n}\,2^{(j-1)\alpha}}$$

$$= 2^\alpha \sum_{\substack{j\geq L\\ 2^j\leq\gamma r_n}} \frac{n^{(\alpha-\beta)/(2\alpha-2\beta)}}{\sqrt{n}\,2^{j(\alpha-\beta)}} = 2^\alpha \sum_{\substack{j\geq L\\ 2^j\leq\gamma r_n}} \frac{1}{2^{j(\alpha-\beta)}} \leq 2^\alpha \sum_{j\geq L}\frac{1}{2^{j(\alpha-\beta)}}.$$

The last series can be made small by taking L big enough since $\alpha>\beta$. Hence, every summand in (1.11) can be made small and thus the theorem is proven. □

Completing the section as well as the chapter, we prove Lemma 1.15. It is mainly based on the Portmanteau Theorem which is stated, for example, in Lemma 2.2, van der Vaart (1998).

Proof of Lemma 1.15. Let

$$\overline{M}_n(u_2,u_2') = M_n(u_2)+M_n'(u_2'), \qquad \overline{M}_n'(u_2,u_2') = M_n'(u_2)+M_n'(u_2'),$$
$$\overline{N}_n(u_2,u_2') = N_n(u_2)+N_n(u_2'), \qquad \overline{N}(u_2,u_2') = N(u_2)+N(u_2').$$

Further, for $B\subset\mathbb{R}^k$ set

$$N_n(B) = \inf_{u_2\in B} N_n(u_2)$$

and similarly for $N(B), M_n(B), M_n'(B)$ as well as $\overline{M}_n(B'), \overline{M}_n'(B'), \overline{N}_n(B'), \overline{N}(B')$ if $B'\subset\mathbb{R}^{2k}$.

We shall show $(\vartheta_n,u_n)\xrightarrow{\mathcal{L}}(u_0,u_0)$, so that from the continuous mapping theorem we deduce that (ϑ_n-u_n) converges to 0 weakly and thus in probability. For the weak convergence of (ϑ_n,u_n) we utilize the Portmanteau Theorem.

Let $A\subset\mathbb{R}^{2k}$ be closed and $\varepsilon>0$. Since both ϑ_n and u_n are stochastically bounded by assumption we can find a compact set $K_0\subset\mathbb{R}^{2k}$ for which $\mathrm{P}((\vartheta_n,u_n)\notin K_0)\leq\varepsilon$ and $\mathrm{P}((u_0,u_0)\notin K_0)\leq\varepsilon$. From (1.9) and the representation of M_n' we have that

$$\overline{M}_n(A\cap K_0) = \overline{M}_n'(A\cap K_0)+o_{\mathrm{P}}(1) = \overline{N}_n(A\cap K_0)+o_{\mathrm{P}}(1)+2\,R_n,$$

and similarly for $\overline{M}_n(K_0)$. Now if $(\vartheta_n,u_n)\in A\cap K_0$, then $\overline{M}_n(A\cap K_0)\leq\overline{M}_n(K_0)$ holds, and by the above this implies $\overline{N}_n(A\cap K_0)\leq\overline{N}_n(K_0)+o_{\mathrm{P}}(1)$. Thus

$$\mathrm{P}((\vartheta_n,u_n)\in A\cap K_0)\leq\mathrm{P}\Big(\overline{N}_n(A\cap K_0)\leq\overline{N}_n(K_0)+o_{\mathrm{P}}(1)\Big) \tag{1.13}$$

is true. The process N_n is asymptotically tight by assumption, hence (N_n,N_n) is asymptotically tight by Lemma 1.4.3, van der Vaart and Wellner (1996). In addition, the convergence of the finite dimensional distributions of (N_n, N_n) is fulfilled as $N_n \rightsquigarrow N$ in $(\ell^\infty(K), \|\cdot\|_K)$ for every compact set $K \subset \mathbb{R}^k$. Thus, Theorem 1.5.4 of van der Vaart and Wellner (1996) yields $(N_n, N_n) \rightsquigarrow (N, N)$ in $(\ell^\infty(K_2), \|\cdot\|_{K_2})$ for any compact set $K_2 \subset \mathbb{R}^{2k}$. Therefore, by the continuous mapping theorem the weak convergence $N_n(u_2) + N_n(u_2') \rightsquigarrow N(u_2) + N(u_2')$ in $\ell^\infty(K_2)$ with respect to the supremum distance follows. Hence, – again due to the continuous mapping theorem – the convergence $(\overline{N}_n(A \cap K_0), \overline{N}_n(K_0)) \xrightarrow{\mathcal{L}} (\overline{N}(A \cap K_0), \overline{N}(K_0))$ holds. Then Slutsky's lemma and the Portmanteau Theorem imply

$$\mathrm{P}\Big(\overline{N}_n(A \cap K_0) \leq \overline{N}_n(K_0) + o_{\mathrm{P}}(1)\Big) \leq \mathrm{P}\Big(\overline{N}(A \cap K_0) \leq \overline{N}(K_0)\Big) + o(1), \tag{1.14}$$

where the $o(1)$-term is a deterministic sequence converging to zero. Since (u_0, u_0) is the unique minimizer of $\overline{N}$ by assumption, on the event $\{(u_0, u_0) \in A^c\}$ the inequality $\overline{N}(u_0, u_0) < \overline{N}(A \cap K_0)$ is fulfilled. If we additionally are on the event $\{\overline{N}(A \cap K_0) \leq \overline{N}(K_0)\}$, we can deduce that $\overline{N}(u_0, u_0) < \overline{N}(K_0)$ must hold, hence $(u_0, u_0) \notin K_0$. This means

$$\mathrm{P}\Big(\overline{N}(A \cap K_0) \leq \overline{N}(K_0)\Big) \leq \mathrm{P}\big((u_0, u_0) \notin K_0\big) + \mathrm{P}\big((u_0, u_0) \in A\big). \tag{1.15}$$

Combining (1.13), (1.14) and (1.15) gives

$$\limsup_{n\to\infty} \mathrm{P}((\vartheta_n, u_n) \in A \cap K_0) \leq \mathrm{P}((u_0, u_0) \in A) + \mathrm{P}((u_0, u_0) \notin K_0).$$

Now, by the choice of K_0, we have $\mathrm{P}((u_0, u_0) \notin K_0) \leq \varepsilon$ and

$$\begin{aligned}\limsup_{n\to\infty} \mathrm{P}((\vartheta_n, u_n) \in A \cap K_0) &\geq \limsup_{n\to\infty} \mathrm{P}((\vartheta_n, u_n) \in A) - \sup_n \mathrm{P}((\vartheta_n, u_n) \notin K_0) \\ &\geq \limsup_{n\to\infty} \mathrm{P}((\vartheta_n, u_n) \in A) - \varepsilon,\end{aligned}$$

so it follows that

$$\limsup_{n\to\infty} \mathrm{P}((\vartheta_n, u_n) \in A) \leq \mathrm{P}((u_0, u_0) \in A) + 2\varepsilon.$$

Since ε was arbitrary, the Portmanteau Theorem yields the desired weak convergence $(\vartheta_n, u_n) \xrightarrow{\mathcal{L}} (u_0, u_0)$. □

Chapter 2

Weak convergence and the hypi-semimetric

In this chapter we reviewed the weak convergence theory for (semi-)metric spaces and applied this to spaces of functions which we equip with the hypi-semimetric on the one hand and the Skorohod M_1- and M_2-distance on the other hand. We showed a general scheme for weak convergence results which is going to be applied in case of the empirical expectile and quantile processes later on.

2.1 Introduction

Many statistics of interest can be written as a functional applied to a stochastic process taking values in the space of bounded functions. It is therefore important to have a weak convergence theory for processes available which behaves well regarding, for example, a potential continuous mapping theorem. The most successful approach was put forward in the 70s by Hoffmann-Jørgensen resulting in today's standard weak convergence theory for stochastic processes. While minimizing the measurability assumptions on the random elements in consideration, it still yields a rich theory applying to a huge diversity of applications.

Regardless of the underlying topology used to define convergence, investigating the weak convergence of a random process is mostly split into two steps. First, the weak convergence of the finite dimensional distributions of the sequence of random processes is shown. Second, a criterion for asymptotic tightness of the sequence is proven, which is often the harder part.

If the considered stochastic processes are indexed by a set of functions, it is possible to connect the tightness criterion to the complexity of the underlying function class by using maximal inequalities. The complexity of the class can be measured, for example, with the bracketing entropy (van der Vaart, 1998, Chapter 19).

Most commonly the space of bounded functions is equipped with the supremum distance (van der Vaart and Wellner, 1996) which is often a fine enough topology. Still, there are examples for which weak convergence with respect to the uniform topology does not hold, for example if the trajectories of the limit process possesses jumps which are not mirrored exactly by the considered sequence. A solution is to use another topology in which convergence is easier to fulfil while ideally collapsing to the uniform topology in regular settings.

This requirement is met by the Skorohod metrics (Skorokhod, 1956) used in the space of real-valued functions in one variable which are right-continuous and have a left-sided limit in every point of their domain. The metrics allow for closeness of two functions, even if jumps do not occur at the same time or not at all in either of them. Skorohod introduced four topologies, called J_1-, M_1-, J_2- and M_2-topology, to attain this in different ways covering many applications where the uniform topology is not suitable. The main focus in the literature is on the J_1-metric (Billingsley, 1999, Chapter 3); few authors also work with the M_1-topology (Avram and Taqqu, 1989; Whitt, 2002).

Turning to functions in several variables with values in $\mathbb{R}^k$, $k > 1$, while maintaining the former smoothness conditions, generalizations of the Skorohod topologies become necessary (Straf, 1969; Neuhaus, 1971; Whitt, 2002). Recently, Bücher et al. (2014) argued that these do not suffice to handle situations as, for example, empirical copula processes. Their solution was to introduce the *hypi-semimetric* which combines ideas originated in optimization theory and can be seen as a "coordinate-free extension of Skorohod M_2-convergence to nonsmooth functions on rahter general domains" (Bücher et al., 2014, page 7). This semimetric is again weaker than the supremum distance while they coincide in certain situations.

The concept of hypi-convergence is a combination of *hy*po- and e*pi*-convergence, explaining the name as an acronym of the latter two. Epi-convergence was introduced in the 60s, see Wijsman (1964), Wijsman (1966), Mosco (1969) and Kall (1986) for an early review, and successfully applied in various fields, see for example the references in Attouch and Wets (1980).

The present chapter proceeds as follows. We explain the Hoffmann-Jørgensen weak convergence theory in (semi-)metric spaces in Section 2.2. An introduction to the concept of entropy with bracketing is given there as well. In Section 2.3 we introduce the hypi-semimetric and the Skorohod M_1- and M_2-distances needed in Chapter 5. Additionally, that section comprises a general scheme applicable to prove a weak convergence result in the hypi-semimetric based on the functional delta-method. All proofs are given in Section 2.4.

2.2 The Hoffmann-Jørgensen weak convergence in (semi-)metric spaces

Here we review the concept of weak convergence of processes due to Hoffmann-Jørgenssen and show a possible extension thereof to semimetric spaces. After this we shortly introduce the bracketing entropy used to measure the complexity of a class of functions.

2.2.1 Weak convergence in metric spaces

When considering maps Y_n, Y from a probability space into the real numbers, the convergence $\mathbb{E}[f(Y_n)] \to \mathbb{E}[f(Y)]$ for all bounded, continuous functions f is equivalent to $Y_n \xrightarrow{\mathcal{L}} Y$. In the basic theory (Billingsley, 1999) the maps Y_n and Y are assumed to be Borel-measurable. When turning to mappings from the probability space into another metric space, the assumption of Borel-measurability of a map is proven to be hard to satisfy – the Borel-σ-field is too big (Chibisov, 1965). This especially occurs if the metric space is non-separable as for example the space of all bounded functions endowed with the supremum metric. The idea was to diminish the measurability assumptions needed for weak convergence by using outer integrals and outer probabilities.

2.1 Definition.

Let $Z : \Omega \to \overline{\mathbb{R}}$ be a map. The *outer integral* of Z is

$$\mathbb{E}^\circ[Z] = \inf \Big\{ \mathbb{E}[U] \mid U : \Omega \to \overline{\mathbb{R}} \text{ measurable}, U \geq Z \text{ and } \min\Big\{\mathbb{E}\Big[U^+\Big], \mathbb{E}\Big[U^-\Big]\Big\} < \infty \Big\}$$

with $x^+ = 0 \vee x$ and $x^- = (-x) \vee 0$. The *outer probability* of an arbitrary set $A \subset \Omega$ is defined by

$$\mathrm{P}^\circ(A) = \inf\{\mathrm{P}(B) \mid A \subset B \text{ and } B \in \mathcal{A}\}.$$

The *inner integral* and *inner probability* are defined accordingly by

$$\mathbb{E}_\circ[Z] = \sup \Big\{ \mathbb{E}[U] \mid U : \Omega \to \overline{\mathbb{R}} \text{ measurable}, U \leq Z \text{ and } \min\Big\{\mathbb{E}\Big[U^+\Big], \mathbb{E}\Big[U^-\Big]\Big\} < \infty \Big\}.$$

and

$$\mathrm{P}_\circ(A) = \sup\{\mathrm{P}(B) \mid A \supset B \text{ and } B \in \mathcal{A}\}.$$

For important properties we refer to Section 1.2, van der Vaart and Wellner (1996). Using outer expectations we define weak convergence for a potentially non-measurable sequence of maps with values in the metric space $(\mathbf{D}, \mathrm{d})$.

2.2 Definition.

Let $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ be a sequence of probability spaces and $Z_n : \Omega_n \to \mathbf{D}$ maps. The sequence Z_n *converges weakly* to a Borel-measurable map $Z : \Omega \to \mathbf{D}$ if $\mathbb{E}^\circ[f(Z_n)] \to \mathbb{E}[f(Z)]$ for every bounded, continuous function $f : \mathbf{D} \to \mathbb{R}$. We write $Z_n \rightsquigarrow Z$ in that case.

At this point we want to stress that the continuity of f depends on the chosen metric d, such that the weak convergence of Z_n depends on d as well.

2.2.2 Weak convergence in semimetric spaces

Let us now turn to a semimetric space $(\mathbf{D}, \mathrm{d})$, so we only assume $\mathrm{d}(x, y) \geq 0$, $\mathrm{d}(x, x) = 0$ and the triangle inequality for elements $x, y \in \mathbf{D}$. Thus, $\mathrm{d}(x, y) = 0$ for $x \neq y$ can happen. In the sequel we use the notation introduced for metric spaces in Section 1.2 for semimetric spaces as well.

When considering convergence of a sequence $x_n \in \mathbf{D}$ to $x \in \mathbf{D}$ with respect to d, a problem is that the limit x must not be unique. If $\mathrm{d}(x_n, x) \to 0$, every element $y \in \mathbf{D}$ with $\mathrm{d}(x, y) = 0$ also fulfils $\mathrm{d}(x_n, y) \to 0$ (see below for an argument). We can fix this flaw by transforming $(\mathbf{D}, \mathrm{d})$ into a metric space as follows. Consider the set $[\mathbf{D}]$ consisting of all equivalence classes $[x]$ of elements $x \in \mathbf{D}$,

$$[\mathbf{D}] = \{[x] \mid x \in \mathbf{D}\}, \text{ where } [x] = \{y \in \mathbf{D} \mid \mathrm{d}(x, y) = 0\}.$$

Now observe that for any $x_1, x_2 \in [x]$ and $y_1, y_2 \in [y]$ it holds that

$$\mathrm{d}(x_1, y_1) \leq \mathrm{d}(x_1, x_2) + \mathrm{d}(x_2, y_2) + \mathrm{d}(y_2, y_1) = \mathrm{d}(x_2, y_2)$$

and vice versa by using the triangle inequality twice. Hence, $\mathrm{d}(x_1, y_1) = \mathrm{d}(x_2, y_2)$ for any representative $x_1, x_2 \in [x]$ and $y_1, y_2 \in [y]$. From this it follows that the map $\mathrm{d}_{[]} : [\mathbf{D}] \times [\mathbf{D}] \to \mathbb{R}$, $([x], [y]) \mapsto \mathrm{d}_{[]}([x], [y])$ with $\mathrm{d}_{[]}([x], [y]) = \mathrm{d}(x, y)$ for arbitrary representatives $x \in [x]$, $y \in [y]$ is well-defined. It is positive, symmetric and fulfils the triangle inequality, as this is the case for d itself. Further, $0 = \mathrm{d}_{[]}([x], [y])$ if and only if $\mathrm{d}(x, y) = 0$ for every representative $x \in [x]$, $y \in [y]$, which means $[x] = [y]$. This reveals that $([\mathbf{D}], \mathrm{d}_{[]})$ is a metric space.

In the following we equip $\mathbf{D}$ with the topology $\mathcal{O}$ of d-open sets. Accordingly, $\mathcal{O}_{[]}$ is the topology on $[\mathbf{D}]$ consisting of the $\mathrm{d}_{[]}$-open sets. Topological terms are understood with respect to these topologies.

We now want to define weak convergence of random elements $Z_n : \Omega_n \to \mathbf{D}$ by looking at the weak convergence of $[Z_n]$ where the latter is defined by virtue of $([\mathbf{D}], \mathrm{d}_{[]})$ being a metric space. For the transition from Z_n to $[Z_n]$ we have to understand the behaviour of $[\cdot]$ and $[\cdot]^{-1}$ on Borel sets. Therefore let $\mathcal{B}(\mathbf{D})$ and $\mathcal{B}([\mathbf{D}])$ be the Borel-σ-algebras on $\mathbf{D}$ and $[\mathbf{D}]$, so they are the smallest σ-algebras containing the respective open sets. The next lemma is crucial to move from Borel laws on $\mathbf{D}$ to Borel laws on $[\mathbf{D}]$ and vice versa.

2.3 Lemma.

i) *The map* $[\cdot] : \mathbf{D} \to [\mathbf{D}]$ *is open, closed and continuous.*

ii) *The closure of* $\{x\} \subset \mathbf{D}$ *is* $[x]$.

iii) *The interior of* $\{x\} \subset \mathbf{D}$ *is either* $\emptyset$ *or* $[x]$.

iv) *The preimage of a set* $[B] \subset [\mathbf{D}]$ *under* $[\cdot]$ *can be written as*

$$[\cdot]^{-1}([B]) = \bigcup_{[x]\in[B]} [x].$$

Thus, $[\cdot]^{-1}([B])$ *is a Borel set.*

v) *If* $x \in B$ *for a set* $B \in \mathcal{B}(\mathbf{D})$, *then* $[x] \subset B$.

Especially, the map $[\cdot] : \mathcal{B}(\mathbf{D}) \to \mathcal{B}([\mathbf{D}])$ *is bijective and* $x \in B$ *if and only if* $[x] \in [B]$ *for* $B \in \mathcal{B}(\mathbf{D})$.

With the next corollary we are able to transform the weak convergence from $\mathbf{D}$ to $[\mathbf{D}]$.

2.4 Corollary.

Let $Z : \Omega \to \mathbf{D}$ *be* $\mathcal{A}$-$\mathcal{B}(\mathbf{D})$*-measurable. Then the map* $[Z] = [\cdot] \circ Z : \Omega \to [\mathbf{D}]$ *is* $\mathcal{A}$-$\mathcal{B}([\mathbf{D}])$*-measurable. Further, every measure* μ *on* $\mathcal{B}(\mathbf{D})$ *induces a measure* $\mu \circ [\cdot]^{-1}$ *on* $\mathcal{B}([\mathbf{D}])$. *Conversely, a measure* $\mu_{[\,]}$ *on* $\mathcal{B}([\mathbf{D}])$ *induces a measure* $\mu_{[\,]} \circ [\cdot]$ *on* $\mathcal{B}(\mathbf{D})$.

This enables the definition of weak convergence in semimetric spaces.

2.5 Definition.

Let $(\Omega_n, \mathcal{A}_n, \mathrm{P}_n)$ be probability spaces. For a sequence of maps $Z_n : \Omega_n \to \mathbf{D}$ we say that Z_n *converges weakly in* $\mathbf{D}$ to an $\mathcal{A}$-$\mathcal{B}(\mathbf{D})$-measurable map $Z : \Omega \to \mathbf{D}$ if $[Z_n]$ converges weakly to $[Z]$ in the sense of Definition 2.2. We write $Z_n \rightsquigarrow Z$ in that case.

A convenient way to deduce weak convergence is to prove the weak convergence of a sequence of random elements, which is easier to handle, and transform the weak convergence to the sequence of interest. For semimetric spaces the common tools, such as the continuous mapping theorem and the delta-method, must be refined in order to address the measurability issues. Formulations of these can be found in Bücher et al. (2014, Appendix B).

2.2.3 Entropy with Bracketing

Next we turn to a more concrete space which is tailored to our needs. Therefore let $(\mathbb{T}, \mathrm{d}_{\mathbb{T}})$ be a compact and separable metric space and collect in $\ell^\infty(\mathbb{T})$ all bounded functions $h : \mathbb{T} \to \mathbb{R}$. The set $\ell^\infty(\mathbb{T})$ can be equipped with the supremum distance $\|\cdot\|_{\mathbb{T}}$; then $(\ell^\infty(\mathbb{T}), \|\cdot\|_{\mathbb{T}})$ is a metric space. We briefly recall conditions for weak convergence of random elements $Z_n : \Omega_n \to \ell^\infty(\mathbb{T})$. Connecting Theorem 1.5.4 and 1.5.7, van der Vaart and Wellner (1996), reveals that for weak convergence of Z_n to a limit process $Z : \Omega \to \ell^\infty(\mathbb{T})$, all of the following assertions need to be true:

i) For every $l \in \mathbb{N}$, $t_1, \dots, t_l \in \mathbb{T}$ the finite dimensional marginals $\big(Z_n(t_1), \dots, Z_n(t_l)\big)$ converge weakly in $\mathbb{R}^l$ to $\big(Z(t_1), \dots, Z(t_l)\big)$;

ii) every $Z_n(t)$ is *asymptotically tight* in $\mathbb{R}$, meaning for every $\varepsilon > 0$ there exists a compact set $K \subset \mathbb{R}$, such that for every $\delta > 0$ it holds that

$$\liminf_{n\to\infty} \mathrm{P}_\circ\big(Z_n(t) \in K^\delta\big) \geq 1 - \varepsilon;$$

iii) for every $\varepsilon, \eta > 0$ there is a $\delta > 0$ which fulfils

$$\limsup_{n\to\infty} \mathrm{P}^\circ\left(\sup_{\mathrm{d}_{\mathbb{T}}(s,t)<\delta} \big|Z_n(s) - Z_n(t)\big| \geq \varepsilon\right) < \eta.$$

The requirement in iii) is called *asymptotically uniformly $\mathrm{d}_{\mathbb{T}}$-equicontinuity in probability.* Note that the term occurring in the outer probability is the *modulus of continuity* of Z_n with respect to the supremum norm and $\mathrm{d}_{\mathbb{T}}$.

Now let $\mathbb{T} = K \subset \mathbb{R}$ be compact and $\mathrm{d}_{\mathbb{T}}$ the Euclidean distance. Further, we fix a family $\mathcal{H} = \{f_t : \mathbb{R} \to \mathbb{R}\}_{t\in K}$ of measurable, bounded functions and a sequence of independent identically distributed real valued random variables $Y, Y_1, \dots Y_n$ with $Y, Y_i \sim F$. We consider weak convergence of the empirical process indexed by $\mathcal{H}$, precisely of the process $\{\mathbb{G}_n[f_t] \,|\, f_t \in \mathcal{H}\}$. This can equally well be seen as a process indexed in K by considering $\{\mathbb{G}_n[f_t] \,|\, t \in K\}$. By the boundedness of f_t, the latter is a random element with values in $\ell^\infty(K)$, so in order to obtain weak convergence of that process, we have to deal with i) – iii) above.

Mostly, i) and ii) can be shown with the aid of a central limit theorem if we suppose that $\mathbb{E}\left[f_t(Y)^2\right] < \infty$. For iii) we give a sufficient condition using the "complexity" of the class $\mathcal{H}$, which is measured by the entropy with bracketing.

2.6 Definition.

Let $l, u : \mathbb{R} \to \mathbb{R}$. A *bracket* $[\![l, u]\!]$is the set of all functions $h : \mathbb{R} \to \mathbb{R}$ with $l \leq h \leq u$. It is called *ε-bracket (with respect to* $\|\cdot\|_{Y,2}$) if $\mathbb{E}\left[\left|u(Y) - l(Y)\right|^2\right] < \varepsilon^2$. An *envelope function* for $\mathcal{H}$ is a function $E : \mathbb{R} \to \mathbb{R}$ with $h \in [\![-E, E]\!]$for all $h \in \mathcal{H}$.
The *bracketing number* $N_{[\,]}(\varepsilon, \mathcal{H}, \|\cdot\|_{Y,2})$ of $\mathcal{H}$ is the minimum number of ε-brackets needed to cover $\mathcal{H}$. Further, define the *entropy with bracketing of* $\mathcal{H}$ as the number $\log\left(N_{[\,]}(\varepsilon, \mathcal{H}, \|\cdot\|_{Y,2})\right)$. The *bracketing integral* $J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})$ is the quantity

$$J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2}) = \int_0^\delta \sqrt{\log\left(N_{[\,]}(\varepsilon, \mathcal{H}, \|\cdot\|_{Y,2})\right)}\, \mathrm{d}\varepsilon.$$

If we can control the size of $J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})$, we can deduce iii) above as follows. By the Markov inequality it holds that

$$\mathrm{P}^\circ\left(\sup_{|s-t|<\delta} \left|\mathbb{G}_n[f_s] - \mathbb{G}_n[f_t]\right| \geq \varepsilon\right) \leq \frac{1}{\varepsilon}\,\mathbb{E}^\circ\left[\sup_{|s-t|<\delta} \left|\mathbb{G}_n[f_s] - \mathbb{G}_n[f_t]\right|\right].$$

The right hand side can be estimated with the aid of Lemma 19.34., van der Vaart (1998), as

$$\mathbb{E}^\circ\left[\sup_{|s-t|<\delta} \left|\mathbb{G}_n[f_s - f_t]\right|\right] \leq C\Big(J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2}) + a(\delta)^{-1}\mathbb{E}\Big[E(Y)^2\, \mathbb{1}\left(E(Y) > \sqrt{n}\, a(\delta)\right)\Big]\Big),$$

where C is some constant not depending on n or δ, $a(\delta)$ is a constant not depending on n and E is a measurable envelope function for $\mathcal{H}$. If E is square integrable, the second summand converges to zero as $n \to \infty$ for every fixed δ. So if $J_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2}) \to 0$ as $\delta \to 0$ we indeed obtain iii) by first letting $n \to \infty$ and then choosing δ small.

Using the supremum distance on $\ell^\infty(\mathbb{T})$ works in many applications. Nevertheless, there are situations in which that metric is too strong, for example if the limit function possesses jumps, which are not matched in the paths of the considered sequence of functions. Especially, a series of continuous functions cannot converge to a discontinuous limit with respect to the supremum distance. For such settings we need other distances.

2.3 The hypi-semimetric and Skorohod M_1/M_2-convergence

Here we recall the hypi-semimetric by Bücher et al. (2014) and two of the Skorohod-topologies, usually referred to as M_1- and M_2-topology. Additionally, we prove a general scheme to obtain weak convergence in the hypi-semimetric, which is based on the delta-method (Bücher et al., 2014, Theorem B.7).

The first properties for the hypi-semimetric concentrates on maps $h : I \to \mathbb{R}$ for an interval $I \subset \mathbb{R}$, which are *càdlàg*: They are right-continuous with existing left-sided limits in every point of I (continu à droite, limites à gauche). Collect in $\mathcal{D}(I)$ all càdlàg-functions and define the acronym *làdcàg* similarly (limites à droite, continu à gauche).

2.3.1 The hypi-semimetric

In order to define the hypi-semimetric we need the epi- and hypographs of $h \in \ell^\infty(K)$ for a compact set $K \subset \mathbb{R}$, which are "the area above and below the function h".

2.7 Definition.

For a function $h \in \ell^\infty(K)$ the *epi*- and *hypographs* are the subsets $\mathrm{epi}(h), \mathrm{hypo}(h) \subset K \times \mathbb{R}$ given by

$$\mathrm{epi}(h) = \big\{(t, y) \in K \times \mathbb{R} \,|\, h(t) \le y\big\} \quad \text{and}$$
$$\mathrm{hypo}(h) = \big\{(t, y) \in K \times \mathbb{R} \,|\, y \le h(t)\big\}.$$

Note that $\mathrm{epi}(h)$ and $\mathrm{hypo}(h)$ are always non-empty for bounded h. The sets are closely related to the semicontinuous hulls of h.

2.8 Definition.

The *lower*- and *upper-semicontinuous hulls* of $h \in \ell^\infty(K)$ are defined by the functions $h_\wedge, h_\vee : K \to \mathbb{R}$,

$$h_\wedge(t) = \lim_{\varepsilon \searrow 0} \inf \big\{h(t') \;|\; t' \in K,\, |t - t'| < \varepsilon\big\} \quad \text{and}$$
$$h_\vee(t) = \lim_{\varepsilon \searrow 0} \sup \big\{h(t') \;|\; t' \in K,\, |t - t'| < \varepsilon\big\}. \tag{2.1}$$

The hulls satisfy $h_\wedge, h_\vee \in \ell^\infty(K)$ as well as $h_\wedge \le h \le h_\vee$. If h is continuous in t, then $h_\vee(t) = h_\wedge(t) = h(t)$. Moreover, it holds that $\mathrm{cl}(\mathrm{epi}(h)) = \mathrm{epi}(h_\wedge)$ and $\mathrm{cl}(\mathrm{hypo}(h)) = \mathrm{hypo}(h_\vee)$ where we equip $K \times \mathbb{R}$ with the Euclidean topology $\mathcal{O}_e$.

In order to get a first impression of how these hulls look like, we present the following.

2.9 Lemma.

Let $h \in \ell^\infty(K)$.

i) If h admits right- and left-sided limits in every $t \in K$, the functions $t \mapsto h(t-)$ and $t \mapsto h(t+)$ do the same. More precisely, the right-sided limit of both $h(t-)$ and $h(t+)$ is $h(t+)$, the left-sided limit of them is $h(t-)$.

ii) Assume h has left- and right-sided limits at every point in K. Then

$$h_\wedge(t) = \min\{h(t-), h(t), h(t+)\} \qquad \text{and}$$
$$h_\vee(t) = \max\{h(t-), h(t), h(t+)\}$$

holds. Especially, if $h(t_0-) \leq h(t_0) \leq h(t_0+)$ or $h(t_0-) \geq h(t_0) \geq h(t_0+)$ for some $t_0 \in K$, then the equalities $(h_\wedge)_\vee(t_0) = h_\vee(t_0)$ and $(h_\vee)_\wedge(t_0) = h_\wedge(t_0)$ are true.

This directly yields a corollary, which enables us to deal with the semicontinuous hulls of products of càdlàg-functions.

2.10 Corollary.

Let $\varphi, h \in \mathcal{D}(I)$ and $t \in I$.

i) If $\big(\varphi(t) - \varphi(t-)\big)\big(h(t) - h(t-)\big) \geq 0$, then it holds that

$$(\varphi\, h)_\wedge(t) = \min\{(\varphi_\wedge\, h_\wedge)(t),\, (\varphi_\vee\, h_\vee)(t)\} \qquad \text{and}$$
$$(\varphi\, h)_\vee(t) = \max\{(\varphi_\wedge\, h_\wedge)(t),\, (\varphi_\vee\, h_\vee)(t)\}.$$

ii) If $h(t) = 0$, then it is satisfied that

$$(\varphi\, h)_\wedge(t) = \min\{\varphi(t-)h(t-),\, 0\} \qquad \text{and} \qquad (\varphi\, h)_\vee(t) = \max\{\varphi(t-)h(t-),\, 0\}.$$

iii) If $h(t-) = 0$, then it is valid that

$$(\varphi\, h)_\wedge(t) = \min\{\varphi(t)h(t),\, 0\} \qquad \text{and} \qquad (\varphi\, h)_\vee(t) = \max\{\varphi(t)h(t),\, 0\}.$$

Although the cases seem restricted, they will be of great help when considering the limit of the empirical quantile process in Chapter 5. The proofs of Lemma 2.9 and Corollary 2.10 are given in Section 2.4.

We now want to define the convergence of functions $h_n \in \ell^\infty(K)$ by looking at the behaviour of their epi- and hypographs. Therefore we review a concept for convergence of sequences of (closed) sets. Let us collect in $\mathbb{F}(K \times \mathbb{R})$ all closed sets defined by $\mathcal{O}_e$, that is all subsets of $K \times \mathbb{R}$ which are closed with respect to the (restricted) Euclidean metric d_e. A well-studied topology on $\mathbb{F}(K \times \mathbb{R})$ is the *Fell-Matheron-topology* or *hit-and-miss-topology* (Ogura, 2007). In general, this topology has to be defined by presenting a base. In our situation, the Fell-Matheron-topology is metrizable, meaning there exists a metric $\mathrm{d}_{\mathbb{F}}$ inducing the topology. For this we observe the following:

- $(K \times \mathbb{R}, \mathrm{d}_e)$ is a locally compact, separable metric space;
- $(K \times \mathbb{R}, \mathcal{O}_e)$ is a *Hausdorff-space*, as for any two $(t, y), (t', y') \in K \times \mathbb{R}$ with $(t, y) \neq (t', y')$ the balls $B_\varepsilon\big((t, y)\big)$ and $B_\varepsilon\big((t', y')\big)$ in $K \times \mathbb{R}$ are disjoint, as long as $2\,\varepsilon < \mathrm{d}_e\big((t, y), (t', y')\big)$;
- $(K \times \mathbb{R}, \mathcal{O}_e)$ is *second-countable*, which means it has a countable base generating the topology (see Exercise 5 in Section 30 of Munkres (2000));
- $(K \times \mathbb{R}, \mathcal{O}_e)$ is *normal*, as for any disjoint, closed sets $A_1, A_2 \subset K \times \mathbb{R}$ the ε-enlargements $A_1^{\varepsilon, K\times\mathbb{R}}$ and $A_2^{\varepsilon, K\times\mathbb{R}}$ are open sets containing A_1 and A_2, respectively, which are disjoint once $2\,\varepsilon < \inf_{a_1 \in A_1} \inf_{a_2 \in A_2} \mathrm{d}_e(a_1, a_2)$.

Using these properties, Theorem 2.5 in Ogura (2007) implies that there is a metric, say $\mathrm{d}_{\mathbb{F}}$, generating the Fell-Matheron-topology on $\mathbb{F}(K \times \mathbb{R})$. A possible choice for $\mathrm{d}_{\mathbb{F}}$ given in that theorem is

$$\mathrm{d}_{\mathbb{F}}(A_1, A_2) = \sup_{x \in K \times \mathbb{R}} \exp(-\mathrm{d}_e(x, 0)) \left|\mathrm{d}_e(x, A_1) \wedge 1 - \mathrm{d}_e(x, A_2) \wedge 1\right|, \quad A_1, A_2 \in \mathbb{F}(K \times \mathbb{R}).$$

Here $\mathrm{d}(x, A) = \inf_{y \in A} \mathrm{d}_e(x, y)$ is the distance from x to A and the conventions $\inf \emptyset = \infty$ and $\sup \emptyset = 0$ are used.

The hypi-convergence now is a combination of epi- and hypo-convergence.

2.11 Definition.

Let $h, h_n \in \ell^\infty(\mathbb{T})$. If $\mathrm{d}_{\mathbb{F}}\big(\mathrm{cl}(\mathrm{epi}(h_n)), \mathrm{cl}(\mathrm{epi}(h))\big) \to 0$, the sequence h_n *epi-converges* to h. It *hypo-converges* to h if $\mathrm{d}_{\mathbb{F}}\big(\mathrm{cl}(\mathrm{hypo}(h_n)), \mathrm{cl}(\mathrm{hypo}(h))\big) \to 0$. We say that h_n *hypi-converges* to $h \in \ell^\infty(\mathbb{T})$ if h_n epi-converges to $h_\wedge$ and hypo-converges to $h_\vee$.

We observe that the epi-convergence of h_n is equivalent to epi-convergence of $h_{n,\wedge}$, as $\mathrm{cl}(\mathrm{epi}(h_n)) = \mathrm{epi}(h_{n,\wedge})$, and analogously for hypo-convergence. Thus, a distance d_{hypi} for the hypi-convergence is given by

$$\mathrm{d}_{hypi}(h, g) = \max\Big\{\mathrm{d}_{\mathbb{F}}\big(\mathrm{epi}(h_\wedge), \mathrm{epi}(g_\wedge)\big), \mathrm{d}_{\mathbb{F}}\big(\mathrm{hypo}(h_\vee), \mathrm{hypo}(g_\vee)\big)\Big\}.$$

Here we can see that symmetry and the triangle inequality for d_{hypi} do hold, as these transform from $\mathrm{d}_{\mathbb{F}}$ and the properties of the maximum (see Lemma 2.24 below). But, by considering for example indicator functions of open and closed subsets of K, we see that there are functions, say g_0 and h_0, which are not equal pointwise but still their lower- and upper-semicontinuous hulls coincide, implying $\mathrm{d}_{hypi}(g_0, h_0) = 0$. Hence, d_{hypi} is only a semimetric and the limit in the former definition must not be unique.

By virtue of Definition 2.5 we can consider the weak convergence of random elements Z_n with values in $(\ell^\infty(K), \mathrm{d}_{hypi})$, which is done in Chapter 5 for the empirical expectile and quantile process.

In order to get a better understanding for the convergence in the hypi-semimetric, we observe that $\mathrm{epi}(h), \mathrm{hypo}(h) \neq \emptyset$ for every $h \in \ell^\infty(K)$, such that we can equally well consider the convergence in the Fell-Matheron-topology only on $\mathbb{F}(K \times \mathbb{R}) \setminus \{\emptyset\}$. By Ogura (2007, Corollary 3.5), the $\mathrm{d}_{\mathbb{F}}$-convergence in that space is equivalent to the *Painlevé-Kuratowski-convergence*, which has a more amenable definition: A sequence of sets $A_n \subset K \times \mathbb{R}$ converges to a set $A \subset K \times \mathbb{R}$ if for every $x \in A$ there exists a sequence $x_n \in A_n$ with $\mathrm{d}_e(x_n, x) \to 0$ and whenever $x_{n_k} \in A_{n_k}$ converges to some $x \in K \times \mathbb{R}$, it must hold that $x \in A$. This leads to the following convenient pointwise criteria for hypi-convergence; see also Molchanov (2005, Chapter 5, Proposition 3.2).

2.12 Proposition (Proposition 2.1, Bücher et al. (2014)).

For $h_n, h \in \ell^\infty(K)$ the following assertions are equivalent:

i) $\mathrm{d}_{hypi}(h_n, h) \to 0$*;*

ii) it holds that

$$\begin{aligned} &\textit{for all } t, t_n \in K \textit{ with } t_n \to t : h_\wedge(t) \leq \liminf_{n\to\infty} h_n(t_n), \\ &\textit{for all } t \in K \textit{ there exist } t_n \in K \textit{ with } \ t_n \to t : h_\wedge(t) = \lim_{n\to\infty} h_n(t_n), \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} &\textit{for all } t, t_n \in K \textit{ with } t_n \to t : \limsup_{n\to\infty} h_n(t_n) \leq h_\vee(t), \\ &\textit{for all } t \in K \textit{ there exist } t_n \in K \textit{ with } \ t_n \to t : \lim_{n\to\infty} h_n(t_n) = h_\vee(t). \end{aligned} \tag{2.3}$$

Bücher et al. (2014) argue that "hypi-convergence is intermediate between uniform convergence and L^p convergence" (Bücher et al., 2014, page 2), compare their Propositions 2.2 and 2.4. Especially, for continuous limits hypi-convergence is equivalent to convergence in the supremum distance. Additionally, convergence of maxima and minima is maintained by the hypi-convergence (Bücher et al., 2014, Proposition 2.4).

On the other hand, we have to be careful when considering convergence of sums of functions as $\mathrm{d}_{hypi}(h_n, h), \mathrm{d}_{hypi}(g_n, g) \to 0$ does not imply $\mathrm{d}_{hypi}(h_n + g_n, f + g) \to 0$. This is mainly because the hull-operators which map a function to its lower- and upper-semicontinuous hulls respectively, do not behave appropriately for sums; for example, $\limsup_n (a_n + b_n) \neq \limsup_n a_n + \limsup_n b_n$. Lemma A.4 in Bücher et al. (2014) shows that in every point of K at least one of the functions h and g has to be continuous to assert the latter equality. The following lemma collects further properties of the hypi-topology and the hull-operator, which are crucial in deriving our main result in Chapter 5. The proof is given in Section 2.4. We denote with $\mathcal{C}(K)$ all continuous functions $h : K \to \mathbb{R}$.

2.13 Lemma.

Let $h, h_n, \varphi_n \in \ell^\infty(K)$ and $\varphi \in \mathcal{C}(K)$.

i) If $\mathrm{d}_{hypi}(\varphi_n, \varphi) \to 0$ and $\mathrm{d}_{hypi}(h_n, h) \to 0$ hold true, then $h_n\,\varphi_n$ hypi-converges to $h\,\varphi$. More precisely $h_n\,\varphi_n$ epi-converges to $(h\,\varphi)_\wedge$ and hypo-converges to $(h\,\varphi)_\vee$ where

$$\begin{aligned}(\varphi\, h)_\wedge &= \varphi\left(h_\wedge \mathbb{1}(\varphi > 0) + h_\vee \mathbb{1}(\varphi < 0)\right) \quad \text{and} \\ (\varphi\, h)_\vee &= \varphi\left(h_\vee \mathbb{1}(\varphi > 0) + h_\wedge \mathbb{1}(\varphi < 0)\right). \end{aligned} \tag{2.4}$$

ii) If $h_n, h \geq c > 0$ and $\mathrm{d}_{hypi}(h_n, h) \to 0$, the convergence $\mathrm{d}_{hypi}(\frac{1}{h_n}, \frac{1}{h}) \to 0$ follows.

2.3.2 A General Framework

Now we introduce a generic way to prove the weak convergence of random elements with respect to the hypi-semimetric using the functional delta-method. In Chapter 5 we discuss that the weak convergence of the empirical expectile and quantile processes are comprised in this scheme.

Fix two (locally) compact separable metric spaces (K_j, d_j), $j = 1, 2$. The next definition is an adapted version of Bücher et al. (2014, Definition B.6).

2.14 Definition.

Let $\mathbf{D}_0, \mathbb{W} \subset \ell^\infty(K_2)$ and $x \in \mathbf{D}_0$.
A map $\zeta : \mathbf{D}_0 \to \ell^\infty(K_1)$ is called *semi-Hadamard differentiable at x tangentially to* $\mathbb{W}$ if there exists a map $\dot\zeta = \dot\zeta_x : \mathbb{W} \to \ell^\infty(K_1)$, called the *semi-derivative of ζ at x*, for which it is valid that for every $w \in \mathbb{W}$, every sequence $t_n \to 0$, $t_n > 0$, and every sequence $w_n \in \ell^\infty(K_2)$ fulfilling $x + t_n\, w_n \in \mathbf{D}_0$ for every $n \in \mathbb{N}$ and $\mathrm{d}_{hypi}(w_n, w) \to 0$

it holds that

$$\mathrm{d}_{hypi}\left(\frac{\zeta(x+t_n\,w_n)-\zeta(x)}{t_n},\,\dot{\zeta}(w)\right)\to 0.$$

Let $\vartheta \in \ell^\infty(K_1)$ be some parameter of interest and let ϑ_n be a sequence of estimators with values in $\ell^\infty(K_1)$. The first part of the general scheme is the following assertion.

2.15 Theorem.

Suppose that there is a map $\zeta : \ell^\infty(K_2) \to \ell^\infty(K_1)$, a parameter $\rho \in \ell^\infty(K_2)$ and a sequence of estimators ρ_n with values in $\ell^\infty(K_2)$, such that

$$a_n\,\mathrm{d}_{hypi}\Big(\vartheta_n - \vartheta - \big(\zeta(\rho_n)-\zeta(\rho)\big), 0\Big) = o_{\mathrm{P}}(1) \tag{2.5}$$

for a sequence $a_n \to \infty$. Further, assume that we have

$$a_n(\rho_n - \rho) \rightsquigarrow Z \tag{2.6}$$

in $(\ell^\infty(K_2), \mathrm{d}_{hypi})$ for a process Z with paths in some $\mathbb{W} \subset \ell^\infty(K_2)$ almost surely, and suppose that ζ is semi-Hadamard differentiable in ρ tangentially to $\mathbb{W}$ with respect to d_{hypi} having semi-derivative $\dot{\zeta}$. Then we have weakly in $(\ell^\infty(K_1), \mathrm{d}_{hypi})$ the convergence

$$a_n(\vartheta_n - \vartheta) \rightsquigarrow \dot{\zeta}(Z). \tag{2.7}$$

We give the proof in Section 2.4. The idea behind the scheme, which already sketches the proof, is as follows. The convergence in (2.6) determines the weak limit of $a_n(\rho_n-\rho)$, which can be transferred to convergence of the sequence $a_n(\zeta(\rho_n)-\zeta(\rho))$ with aid of the functional delta-method (Bücher et al., 2014, Theorem B.7). But as $a_n(\zeta(\rho_n)-\zeta(\rho))$ and $a_n(\vartheta_n-\vartheta)$ are close in probability by (2.5), this also yields the weak limit of the latter sequence.

The former theorem already suffices to handle the empirical quantile process in Chapter 5. For the analogue for expectiles we consider the following lemma, which may be useful to verify (2.6). The proof can also be found in Section 2.4.

2.16 Lemma.

Let $\xi_0 : \ell^\infty(K_1) \to \ell^\infty(K_2)$ be some functional and suppose that for a sequence of functionals $\xi_n : \ell^\infty(K_1) \to \ell^\infty(K_2)$, such that $\xi_n(\vartheta)$ is a random element in $\ell^\infty(K_2)$, *the convergence*

$$a_n\big(\xi_n(\vartheta) - \xi_0(\vartheta)\big) \rightsquigarrow Z \tag{2.8}$$

in $(\ell^\infty(K_2), \mathrm{d}_{hypi})$ *is true for some process* Z *and* $a_n \to \infty$. *In addition, suppose that*

$$\sup_{\mathrm{d}_{hypi}(\varphi_n, 0) \leq \delta_n} a_n \, \mathrm{d}_{hypi}\big(\xi_n(\vartheta) - \xi_0(\vartheta + \varphi_n),\, 0\big) \tag{2.9}$$

converges to zero in outer probability for any sequence $\delta_n \searrow 0$. *Then the consistency* $\mathrm{d}_{hypi}(\widehat{\vartheta}_n - \vartheta,\, 0) = o_{\mathrm{P}}(1)$ *implies* (2.6) *for* $\rho_n = \xi_0(\vartheta_n)$ *and* $\rho = \xi_0(\vartheta)$.

Here, we first transform the variables ϑ_n with some ξ_0 and want to obtain (2.6) for the transformed process $a_n\big(\xi_0(\vartheta_n) - \xi_0(\vartheta)\big)$. The condition in (2.9) ensures that we can equally well consider the process in (2.8), whose weak limit is given. Thus, (2.6) is fulfilled and we can apply Theorem 2.15 to deduce the weak convergence of $a_n(\vartheta_n - \vartheta)$, provided $\zeta = \xi_0^{\mathrm{Inv}}$ exists and is semi-Hadamard differentiable.

2.3.3 Skorohod M_1- and M_2-Convergence

Let us now choose an interval $I \subset \mathbb{R}$ and consider the set $\mathcal{D}(I)$. For convenience, we formulate the theory for $I = [0, 1]$ only. Above we argued that using convergence in the supremum distance is not appropriate for limiting functions with jumps unmatched in the considered sequence. Skorokhod (1956) introduced four topologies on $\mathcal{D}([0, 1])$, called J_1-, J_2-, M_1- and M_2-topology, in which convergence in such situations is possible. The topologies provide different ways how a jump in the limit function can be approximated by the converging sequence; the hierarchy among the topologies can be found in Whitt (2002, Chapter 11.5.2).

In this section we discuss the M_1- and M_2-topologies. Molchanov (2005) states that convergence in the M_2-topology is equivalent to the hypi-convergence – at least in our situations (Molchanov, 2005, page 377). As no proof for the equivalence of the topologies is given in the former reference, we investigate this in more detail below. Observe that this discussion restricts the hypi-topology to $\mathcal{D}([0, 1])$ as well, but, as the applications in Chapter 5 work with càdlàg-processes only, we still obtain yet another interpretation for the hypi-convergence in our setting. We will also touch on the M_1-metric because we use that metric as a tool to show M_2-convergence in Chapter 5: The M_1-topology is a finer topology than the one generated by the M_2-distance.

We adopt the approach of Pomarede (1976) who was able to unify the definition of the Skorohod topologies by considering (un-)completed graphs and parametric representations thereof.

2.17 Definition.

Let $h \in \mathcal{D}([0,1])$. The *completed graph* $\Gamma_h \subset [0,1] \times \mathbb{R}$ is the set

$$\Gamma_h = \{(t,y) \in [0,1] \times \mathbb{R} \,|\, y \in [h(t-), h(t)]\}.$$

We define the order relation $\leq$ on Γ_h by setting $(t_1,y_1) \leq (t_2,y_2)$ if either $t_1 < t_2$ or $t_1 = t_2$ and $|h(t_1-) - y_1| \leq |h(t_1-) - y_2|$ for elements $(t_1,y_1),(t_2,y_2) \in \Gamma_h$.

The completed graph therefore is a connected curve in $[0,1] \times \mathbb{R}$, where, heuristically, jumps of h are joined with a vertical line. The order relation, which induces a total order on Γ_h, is needed for the M_1-parametric representations of a function h.

2.18 Definition.

Let $h \in \mathcal{D}([0,1])$.

i) A map $(r,u) : [0,1] \to \Gamma_h$ is called *(strong) M_1-parametric representation of h* if it is surjective, continuous, and non-decreasing with respect to the order relation $\leq$ in Definition 2.17. Collect in $\prod_{s,1}(h)$ all M_1-parametric representations of h.

ii) A map $(r,u) : [0,1] \to \Gamma_h$ is called *(strong) M_2-parametric representation of h* if it is surjective, continuous and if $r : [0,1] \to [0,1]$ is increasing. Denote by $\prod_{s,2}(h)$ the set of M_2-parametric representations of h.

Both parametric representations visit all of Γ_h, where r governs "how fast we move" and u the "altitude". So we can view r as the *time part* and u as the *spatial part* of a parametric representation. The difference in i) and ii) becomes clear when considering a function h jumping up in t_0.After hitting $h(t_0-)$, the spatial part $u_1(s)$ in an M_1-parametric representation has to move directly to $h(t_0)$ as s increases which is due to the required monotonicity in i). In contrast, the spatial part $u_2(s)$ in an M_2-parametric representation can move "back and forth" on the segment $[h(t_0-), h(t_0)]$ as no monotonicity is needed in ii). Using these representations, we define the M_1- and M_2-distance.

2.19 Definition.

Given $h_1, h_2 \in \mathcal{D}([0,1])$, define their M_1-*distance* by

$$\mathrm{d}_{s,1}(h_1,h_2) = \inf_{\substack{(r_j,u_j) \in \prod_{s,1}(h_j) \\ j=1,2}} \max\{\|r_1 - r_2\|_{[0,1]},\ \|u_1 - u_2\|_{[0,1]}\}. \tag{2.10}$$

As ordinary, a sequence of functions $h_n \in \mathcal{D}([0,1])$ converges in the M_1-topology to $h \in \mathcal{D}([0,1])$ if $\mathrm{d}_{s,1}(h_n, h) \to 0$ holds.

2.20 Definition.

Given $h_1, h_2 \in \mathcal{D}([0,1])$, the M_2-*distance* between h_1 and h_2 is

$$\mathrm{d}_{s,2}(h_1, h_2) = \inf_{\substack{(r_j,u_j)\in\prod_{s,2}(h_j)\\ j=1,2}} \max\left\{\|r_1 - r_2\|_{[0,1]},\ \|u_1 - u_2\|_{[0,1]}\right\}. \tag{2.11}$$

As above, a sequence of functions $h_n \in \mathcal{D}([0,1])$ converges in the M_2-topology to $h \in \mathcal{D}([0,1])$ if $\mathrm{d}_{s,2}(h_n, h) \to 0$ holds.

The M_2-distance originally is defined as the Hausdorff-distance between the completed graphs. This is equivalent to the above definition, see Whitt (2002, Theorem 12.11.1).

Since $\prod_{s,1} \subset \prod_{s,2}$, the infimum in (2.10) is smaller than the one in (2.11), showing that M_1-convergence implies M_2-convergence. Note that the M_1-distance is a metric (Whitt, 2002, Theorem 12.3.1) whereas the M_2-distance is just a semimetric (Whitt, 2002, Example 12.11.1).

2.21 Remark.
Let $g_1 \in \mathcal{D}([0,1])$ have a unique jump in $t_0 \in [0,1]$ with $g_1(t_0) - g_1(t_0-) = \delta > 0$. A function $g_2 \in \mathcal{D}([0,1])$ can be close to g_1 in the M_2-sense in various ways. For example,

i) it can mirror the jump of g_1 nearly exactly, in the sense that g_2 has a jump in a neighbourhood of t_0 with height similar to δ;

ii) the function g_2 could have a number of jumps around t_0, whose total height is nearly δ;

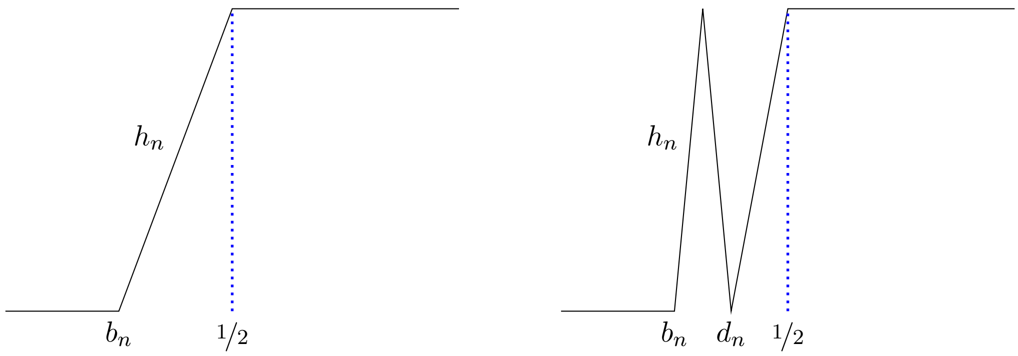
iii) g_2 interpolates between, for example, $g_1(t_0 - \varepsilon)$ and $g_1(t_0)$ for some small $\varepsilon > 0$.

Especially, a continuous function can be close to a discontinuous one.

Note that in iii) the interpolation does not have to be monotonic. This means that, when choosing $t_n \nearrow t_0$, it can happen that $g_2(t_n)$ varies in $[g_1(t_0-), g_1(t_0)]$, giving an oscillating behaviour of g_2 around t_0. This is in contrast to g_2 being close to g_1 in the M_1-sense, where such oscillations are not possible due to the monotonicity in the spatial part of the parametric representations.

For an example and better intuition see Figure 2.1, which is a reproduction of Figure 11.2 in Chapter 11.2 on page 461, Whitt (2002). $\diamond$

Interesting for our application is the following connection between the M_2-distance and the hypi-semimetric, which is proven in Section 2.4.



(a) Close in the M_1-sense. **(b)** Close in the M_2-, but not in the M_1-sense.

Figure 2.1: The black lines depict functions $h_n \in \mathcal{D}([0,1])$, which are close to the indicator $h(x) = \mathbb{1}(x \in [1/2, 1])$ in (a) the M_1-topology and (b) the M_2-sense. We chose $b_n = \frac{1}{2} - \frac{2}{n}$ and $d_n = \frac{1}{2} - \frac{1}{n}$ and depicted h_8.

2.22 Theorem.

Let $h_n, h \in \mathcal{D}([0,1])$. Then $\mathrm{d}_{hypi}(h_n, h) \to 0$ if and only if $\mathrm{d}_{s,2}(h_n, h) \to 0$.

In Chapter 5 we show the semi-Hadamard differentiability of a functional with respect to the hypi-semimetric, where we use Lemma 2.13 to reduce the complexity of the considered sequence. In order to obtain this result for M_2-convergence as well – without using the equivalence of the topologies – we formulate an M_2-version of Lemma 2.13. The proof is also deferred to Section 2.4.

2.23 Lemma.

Let $h, h_n, \varphi_n \in \mathcal{D}([0,1])$ and $\varphi \in \mathcal{C}([0,1])$.

i) If $\mathrm{d}_{s,2}(\varphi_n, \varphi) \to 0$ and $\mathrm{d}_{s,2}(h_n, h) \to 0$ hold true, then $h_n \varphi_n$ converges to $h \varphi$ in the M_2-sense.

ii) If h_n, $h \geq c > 0$, then $\mathrm{d}_{s,2}(h_n, h) \to 0$ implies $\mathrm{d}_{s,2}(1/h_n, 1/h) \to 0$.

Concluding this section we want to stress the M_2-continuity of addition. For functions $h_n, \varphi_n, h, \varphi \in \mathcal{D}([0,1])$ Theorem 12.11.6 in Whitt (2002) states that $h_n + \varphi_n$ converges in the M_2-topology to $h + \varphi$, provided $\mathrm{d}_{s,2}(h_n, h) \to 0$, $\mathrm{d}_{s,2}(\varphi_n, \varphi) \to 0$, and for every $t \in [0,1]$ it holds that $\big(h(t) - h(t-)\big)\big(\varphi(t) - \varphi(t-)\big) \geq 0$. The latter assertion guarantees that the jumps of h and φ do not have opposite signs. Due to Theorem 2.22, the statement holds in the hypi-semimetric as well, generalizing Bücher et al. (2014, Lemma A.4).

2.4 Proofs

Here we give the proofs for the present chapter.

2.4.1 Proofs for Section 2.2.2

We start with the proof of Lemma 2.3, dealing with the properties of building equivalence classes in a semimetric space.

Proof of Lemma 2.3. To prove i) let $U \subset \mathbf{D}$ be open and $x \in U$, so $[x] \in [U]$. Then choose $\varepsilon > 0$, such that $\mathrm{d}(x, y) < \varepsilon$ for $y \in \mathbf{D}$ implies $y \in U$. Now take any $[y] \in [\mathbf{D}]$ with $\mathrm{d}_{[]}([x], [y]) < \varepsilon$, which is equivalent to $\mathrm{d}(x, y) < \varepsilon$ for any representative $y \in [y]$. Hence, $y \in U$ is true, which yields $[y] \in [U]$. So the map $[\cdot]$ sends open sets to open sets, which means that $[\cdot]$ is open. For closed sets this works accordingly.

To obtain the continuity in i) note that $\mathrm{d}(x_n, x) < \varepsilon$ holds for $x_n, x \in \mathbf{D}$ if and only if $\mathrm{d}_{[]}([x_n], [x]) < \varepsilon$ for $[x_n], [x] \in [\mathbf{D}]$.

In order to show ii) let $x \in \mathbf{D}$. For any $y \in \mathrm{cl}(\{x\})$ there must be a sequence $x_n \in \{x\}$, such that $\mathrm{d}(x_n, y) \to 0$. As $\{x\}$ has only one element, $x_n = x$ holds, such that $0 = \lim_n \mathrm{d}(x_n, y) = \mathrm{d}(x, y)$. Hence, $y \in [x]$ is true. On the other hand, for every $y \in [x]$, the sequence $x_n = x$ converges to y with respect to d, which means $y \in \mathrm{cl}(\{x\})$.

Let us turn to iii) and choose $x \in \mathbf{D}$ again. The only possible subsets of $\{x\}$ are $\emptyset$ and $\{x\}$ itself. If $\mathrm{int}(\{x\}) = \{x\}$, by definition there must be an $\varepsilon > 0$, such that $\{x\} \supset B_\varepsilon^{\mathbf{D}}(x) \supset [x]$. Thus, $[x] = \{x\} = \mathrm{int}(\{x\})$ holds, which means that either $\emptyset$ or $[x]$ equals $\mathrm{int}(\{x\})$.

We now prove iv). Choose any $[B] \subset [\mathbf{D}]$ and let $y \in \mathbf{D}$ with $[y] \in [B]$. Then $y \in [y] \subset \bigcup_{[x] \in [B]} [x]$, showing $[\cdot]^{-1}([B]) \subset \bigcup_{[x] \in [B]} [x]$. Conversely, let $y \in \bigcup_{[x] \in [B]} [x]$. Thus, there is an $[x] \in [B]$ with $y \in [x]$, which means $[y] = [x] \in [B]$. Hence, $y \in [\cdot]^{-1}([B])$, such that $\bigcup_{[x] \in [B]} [x] \subset [\cdot]^{-1}([B])$. Last, $\bigcup_{[x] \in [B]} [x]$ is a Borel set as a union of closed sets by ii).

For v) it suffices to show the property for a closed set $B \subset \mathbf{D}$, as these sets form an $\cap$-stable generator of $\mathcal{B}(\mathbf{D})$. So let $B \subset \mathbf{D}$ be closed and $x \in B$. As $\{x\} \subset B$, $\mathrm{cl}(\{x\}) \subset \mathrm{cl}(B) = B$ follows. Assertion ii) thus implies $[x] \subset B$ as asserted.

To reveal that $[\cdot] : \mathcal{B}(\mathbf{D}) \to \mathcal{B}([\mathbf{D}])$ is a bijection, note first that for every $[B] \in \mathcal{B}([\mathbf{D}])$, it holds that $B' = [\cdot]^{-1}([B]) \in \mathcal{B}(\mathbf{D})$ by iv). But, $[B'] = [B]$ holds by definition of the preimage, so $[\cdot]$ is surjective. Next, let $B_1, B_2 \in \mathcal{B}(\mathbf{D})$ with $[B_1] = [B_2]$. If $B_1 = \emptyset$ or $B_2 = \emptyset$, $\emptyset = [B_1] = [B_2]$ follows. Without loss of generality assume $B_1 = \emptyset$. If B_2 had an element $x \in B_2$, then $[x] \in [B_2]$, a contradiction. Thus, $B_2 = \emptyset$ holds as well. Now assume $B_1 \neq \emptyset$ and let $x \in B_1$. Assertion v) implies $[x] \in [B_1] = [B_2]$, such that there is a $y \in B_2$ with $[x] = [y]$. Using v) again, it follows that $x \in [y] \subset B_2$, which shows $B_1 \subset B_2$. Changing the role of B_1 and B_2 implies $B_1 = B_2$ and thus injectiveness of the map $[\cdot]$.

Last let $B \in \mathcal{B}(\mathbf{D})$. We want to show that $x \in B$ if and only if $[x] \in [B]$. The "only if" follows by definition of $[B] = \{[x] \,|\, x \in B\}$. For the "if"-part note that with v) we can write

$$B = \bigcup_{y \in B} \{y\} = \bigcup_{y \in B} [y] = \bigcup_{[y] \in [B]} [y].$$

As $[x] \in [B]$, this means $[x] \subset B$, especially $x \in B$ as asserted. □

2.4.2 Proofs for Section 2.3.1

Next we prove Lemma 2.9, which gave a first impression of the construction of semicontinuous hulls.

Proof of Lemma 2.9. Ad i): This is Lemma C.5, Bücher et al. (2014); we give a proof here for convenience. We show $\lim_{t' \searrow t} h(t'-) = h(t+)$. First we consider the expression $\liminf_{t' \searrow t} h(t'-)$. Observe that

$$\liminf_{t' \searrow t} h(t'-) = \lim_{\varepsilon \searrow 0} \inf_{t' \in (t, t+\varepsilon)} \lim_{t'' \nearrow t'} h(t'') = \lim_{\varepsilon \searrow 0} \inf_{t' \in (t, t+\varepsilon)} \lim_{\delta \searrow 0} \inf_{t'' \in (t'-\delta, t')} h(t'').$$

Therefore choose any $\varepsilon > 0$ and $t' \in (t, t + \varepsilon)$. Then for some small $\delta > 0$ it holds that $(t' - \delta, t') \subset (t, t + \varepsilon)$ and thus

$$\lim_{\delta \searrow 0} \inf_{t'' \in (t'-\delta, t')} h(t'') \geq \inf_{t'' \in (t, t+\varepsilon)} h(t'')$$

is valid. Taking the infimum over $t' \in (t, t + \varepsilon)$ and then letting $\varepsilon \searrow 0$ yields

$$\lim_{\varepsilon \searrow 0} \inf_{t' \in (t, t+\varepsilon)} \lim_{\delta \searrow 0} \inf_{t'' \in (t'-\delta, t')} h(t'') \geq \lim_{\varepsilon \searrow 0} \inf_{t'' \in (t, t+\varepsilon)} h(t'') = \lim_{t'' \searrow t} h(t'') = h(t+),$$

as h has a right-sided limit in t. This means

$$\liminf_{t' \searrow t} h(t'-) \geq h(t+).$$

Similarly we deduce that

$$\limsup_{t' \searrow t} h(t'-) \leq h(t+),$$

hence we obtain

$$\lim_{t' \searrow t} h(t'-) = h(t+)$$

as asserted. The remaining assertions are proven analogously.

Ad ii): The proof of Lemma C.6, Bücher et al. (2014), shows that for a function, which admits right- and left-sided limits in every point, the supremum over a shrinking neighbourhood around a point t converges to the maximum of the three points $h(t-)$, $h(t)$ and $h(t+)$. The analogous statement holds for the infimum, which is the first part of ii).

From ii) before, the maps $t \mapsto h(t-)$ and $t \mapsto h(t+)$ both have a right-sided limit equal to $h(t+)$ and a left-sided limit equal to $h(t-)$, hence this is also true for the functions $h_\vee(t) = \max\{h(t-), h(t), h(t+)\}$ and $h_\wedge(t) = \min\{h(t-), h(t), h(t+)\}$. From the above argument we obtain

$$(h_\vee)_\wedge(t) = \min\left\{h(t-), h(t+), \max\{h(t-), h(t), h(t+)\}\right\} = h(t-) \wedge h(t+) \quad \text{and}$$
$$(h_\wedge)_\vee(t) = \max\left\{h(t-), h(t+), \min\{h(t-), h(t), h(t+)\}\right\} = h(t-) \vee h(t+).$$

If $h(t_0-) \le h(t_0) \le h(t_0+)$ or $h(t_0-) \ge h(t_0) \ge h(t_0+)$, we deduce

$$h_\vee(t_0) = h(t-) \vee h(t+) = (h_\wedge)_\vee(t_0) \quad \text{and} \quad h_\wedge(t_0) = h(t-) \wedge h(t+) = (h_\vee)_\wedge(t_0). \quad \square$$

Proof of Corollary 2.10. This is an application of Lemma 2.9, ii), where the expressions simplify due to the càdlàg-property of the considered maps. $\square$

For the proof of Lemma 2.13 we shall require the following basic relations between lim sup and lim inf.

2.24 Lemma.

Let $(a_n)_n$, $(b_n)_n$ be bounded sequences.

i) It holds that

$$\liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \ge \liminf_{n\to\infty}(a_n + b_n) \ge \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n$$

and

$$\limsup_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \le \limsup_{n\to\infty}(a_n + b_n) \le \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n.$$

ii) If $a_n \ge c > 0$ for all $n \in \mathbb{N}$, then

$$\liminf_n \frac{1}{a_n} = \frac{1}{\limsup_n a_n}$$

is valid; a similar equation holds for $\limsup_n \frac{1}{a_n}$.

iii) Provided $(a_n)_n$ is convergent with limit $a \in \mathbb{R}$, the equalities

$$\liminf_n a_n\, b_n = \liminf_n a\, b_n \quad \text{and} \quad \limsup_n a_n\, b_n = \limsup_n a\, b_n$$

are true.

Proof. The pair of inequalities in i) follows from

$$a_k + \sup_{l \geq n} b_l \geq a_k + b_k \geq a_k + \inf_{l \geq n} b_l$$

for any $n \in \mathbb{N}$, $k \geq n$, by applying "$\inf_{k \geq n}$" across the display and taking the limit $n \to \infty$. The second pair of inequalities in i) follows similarly.

For part ii) note that

$$\inf_{l \geq n} \frac{1}{a_l} = \frac{1}{\sup_{l \geq n} a_l}$$

is valid. Taking $n \to \infty$ yields the asserted equality.

To obtain iii) observe that by i) it is true that

$$\liminf_n (a_n\, b_n - a\, b_n) + \liminf_n a\, b_n \leq \liminf_n a_n\, b_n \leq \limsup_n (a_n\, b_n - a\, b_n) + \liminf_n a\, b_n,$$
$$\liminf_n (a_n\, b_n - a\, b_n) + \limsup_n a\, b_n \leq \limsup_n a_n\, b_n \leq \limsup_n (a_n\, b_n - a\, b_n) + \limsup_n a\, b_n.$$

But $\left|a\, b_n - a_n\, b_n\right| \leq \sup_{l \in \mathbb{N}} |b_l| \left|a - a_n\right| \to 0$ holds, which implies the stated result. $\square$

This helps in the proof of the properties of hypi-convergence.

Proof of Lemma 2.13. From the definition in (2.1), for a function $h \in \ell^\infty(K)$ the lower semi-continuous hull $h_\wedge$ at $t \in K$ is characterized by the following conditions:

$$\begin{aligned} &\text{For any sequence } t_n \to t \quad \text{we have } \liminf_{n\to\infty} h(t_n) \geq h_\wedge(t) \quad \text{and} \\ &\text{there is a sequence } t'_n \to t \quad \text{for which } \lim_{n\to\infty} h(t'_n) = h_\wedge(t), \end{aligned} \tag{2.12}$$

and similarly for $h_\vee$.

Ad i): By continuity of φ we have $\varphi(t_n) \to \varphi(t)$ for any sequence $t_n \to t$. The statement (2.4) now follows immediately using (2.12) and Lemma 2.24 and noting that for $\varphi(t) < 0$

$$\liminf_{n\to\infty} \varphi(t) h(t_n) = \varphi(t) \limsup_{n\to\infty} h(t_n) \quad \text{and} \quad \limsup_{n\to\infty} \varphi(t) h(t_n) = \varphi(t) \liminf_{n\to\infty} h(t_n)$$

are valid. Further, by continuity of φ, the hypi-convergence of φ_n to φ actually implies the uniform convergence. Therefore, for any $t_n \to t$ we have that $\varphi_n(t_n) \to \varphi(t)$. Using the pointwise criteria (2.2) and (2.3) for hypi-convergence, Lemma 2.24 and (2.4) we obtain the asserted convergence $\varphi_n\, h_n \to \varphi\, h$ with respect to the hypi-semimetric.

Ad ii): From ii) in Lemma 2.24 and (2.12) we obtain $(1/h)_\wedge = 1/h_\vee$ and $(1/h)_\vee = 1/h_\wedge$. The hypi-convergence of the sequence $1/h_n$ to these hulls follows similarly from Lemma 2.24, ii), and the pointwise criteria (2.2) and (2.3) for hypi-convergence. $\square$

2.4.3 Proofs for Section 2.3.2

Next we prove the general scheme provided in Theorem 2.15 and Lemma 2.16.

Proof of Theorem 2.15. We can rewrite

$$a_n(\vartheta_n - \vartheta) = a_n(\zeta(\rho_n) - \zeta(\rho)) + a_n[\vartheta_n - \vartheta - (\zeta(\rho_n) - \zeta(\rho))].$$

The term in angle brackets on the right hand side converges to the map $t \mapsto 0$ with respect to d_{hypi} in probability by assumption. Since this limit is continuous and

$$a_n(\zeta(\rho_n) - \zeta(\rho)) \rightsquigarrow \dot{\zeta}(Z)$$

weakly in $(\ell^\infty(K_1), \mathrm{d}_{hypi})$ by Theorem B.7, Bücher et al. (2014), the conclusion (2.7) follows from Lemma A.4 in Bücher et al. (2014). $\square$

Lemma 2.16 then asserted the weak convergence of $a_n(\rho_n - \rho)$ to Z for ρ_n and ρ being itself transformations of ϑ_n and ϑ.

Proof of Lemma 2.16. Write

$$a_n(\rho_n - \rho) = a_n(\xi_n(\vartheta) - \xi_0(\vartheta)) + a_n[\xi_0(\vartheta_n) - \xi_n(\vartheta)]. \tag{2.13}$$

In order to show convergence in probability to 0 of the second term on the right hand side, given $\delta_n, \varepsilon > 0$ we estimate

$$\begin{aligned}
&\mathrm{P}\Big(a_n\, \mathrm{d}_{hypi}(\xi_0(\vartheta_n) - \xi_n(\vartheta), 0) \geq \varepsilon\Big) \\
\leq\, &\mathrm{P}\Big(a_n\, \mathrm{d}_{hypi}(\xi_0(\vartheta_n) - \xi_n(\vartheta), 0) \geq \varepsilon, \mathrm{d}_{hypi}(\vartheta_n - \vartheta, 0) \leq \delta_n\Big) + \mathrm{P}\Big(\mathrm{d}_{hypi}(\vartheta_n - \vartheta, 0) > \delta_n\Big) \\
\leq &\mathrm{P}^\circ\Big(\sup_{\mathrm{d}_{hypi}(\varphi_n, 0) \leq \delta_n} a_n\, \mathrm{d}_{hypi}(\xi_0(\vartheta + \varphi_n) - \xi_n(\vartheta),\, 0) \geq \varepsilon\Big) + \mathrm{P}\Big(\mathrm{d}_{hypi}(\vartheta_n - \vartheta, 0) > \delta_n\Big).
\end{aligned} \tag{2.14}$$

The assumption $\mathrm{d}_{hypi}(\vartheta_n - \vartheta,\, 0) = o_{\mathrm{P}}(1)$ implies that for a sequence $\delta_n \searrow 0$ sufficiently slowly we have as $n \to \infty$ that

$$\mathrm{P}\Big(\mathrm{d}_{hypi}(\vartheta_n - \vartheta, 0) > \delta_n\Big) \to 0.$$

From assumption (2.9), the first term in (2.14) also tends to zero so that we deduce $a_n\, \mathrm{d}_{hypi}(\xi_0(\vartheta_n) - \xi_n(\vartheta), 0) = o_{\mathrm{P}}(1)$. The conclusion follows from (2.8), (2.13), and Slutzky's Lemma. $\square$

2.4.4 Proofs of Section 2.3.3

Here we turn to the proofs concerning the M_2-topology on $\mathcal{D}([0,1])$. To begin with, we want to prove the equivalence of the M_2- and hypi-topology, for which we need the following characterisations of these types of convergence. The first states that convergence in the M_2-sense is equivalent to convergence of certain infima and suprema. For this let $\mathrm{Disc}(\varphi)$ denote the set of all discontinuities of a function $\varphi \in \mathcal{D}([0,1])$.

2.25 Theorem (Theorem 12.11.7, Whitt (2002)).

There is convergence $\mathrm{d}_{s,2}(\varphi_n, \varphi) \to 0$ *if and only if*

$$\sup_{t\in[t_1,t_2]} \varphi_n(t) \to \sup_{t\in[t_1,t_2]} \varphi(t) \quad \textit{and}$$
$$\inf_{t\in[t_1,t_2]} \varphi_n(t) \to \inf_{t\in[t_1,t_2]} \varphi(t)$$

hold true for all points $t_1, t_2 \in \{1\} \cup \mathrm{Disc}(\varphi)^c$, $t_1 < t_2$.

Second, hypi-convergence is related to convergence of infima and suprema as follows.

2.26 Proposition (Proposition 5.3.2, Molchanov (2005)).

Let $\varphi_n, \varphi \in \mathcal{D}([0,1])$. *The following statements are equivalent.*

i) φ_n *epi-converges to* $\varphi_\wedge$.

ii) For every compact set $K \subset [0,1]$ *and every open set* $G \subset [0,1]$ *it holds that*

$$\liminf_n \inf_{t\in K} \varphi_n(t) \geq \inf_{t\in K} \varphi_\wedge(t) \quad \textit{and}$$
$$\limsup_n \inf_{t\in G} \varphi_n(t) \leq \inf_{t\in G} \varphi_\wedge(t).$$

In addition, there is equivalence between the following assertions.

i) φ_n *hypo-converges to* $\varphi_\vee$.

ii) $-\varphi_n$ *epi-converges to* $-(\varphi_\vee) = (-\varphi)_\wedge$.

iii) For every compact set $K \subset [0,1]$ *and every open set* $G \subset [0,1]$ *it holds that*

$$\limsup_n \sup_{t\in K} \varphi_n(t) \leq \sup_{t\in K} \varphi_\vee(t) \quad \textit{and}$$
$$\liminf_n \sup_{t\in G} \varphi_n(t) \geq \sup_{t\in G} \varphi_\vee(t).$$

These equivalences are now used to prove Theorem 2.22.

Proof of Theorem 2.22 Let $\mathrm{d}_{hypi}(\varphi_n, \varphi) \to 0$. Then for any $t_1, t_2 \in \{1\} \cup \mathrm{Disc}(\varphi)^c$, $t_1 < t_2$, it holds that

$$\limsup_n \sup_{t \in [t_1, t_2]} \varphi_n(t) \leq \sup_{t \in [t_1, t_2]} \varphi_\vee(t)$$

by Proposition 2.26. Now we distinguish $t_2 \in \mathrm{Disc}(\varphi)^c$ and $t_2 = 1$.

If $t_2 \in \mathrm{Disc}(\varphi)^c$, then

$$\liminf_n \sup_{t \in (t_1, t_2)} \varphi_n(t) \geq \sup_{t \in (t_1, t_2)} \varphi_\vee(t) = \sup_{t \in [t_1, t_2]} \varphi_\vee(t)$$

follows by Proposition 2.26 and continuity of φ in t_1 and t_2.

Last let $t_2 = 1$. Then, using Proposition 2.12, there is a sequence $t_n \in [t_1, 1]$ with $t_n \to 1$ and $\liminf_n \varphi_n(t_n) = \varphi_\vee(1)$. Thus, it holds that

$$\liminf_n \sup_{t \in [t_1, 1]} \varphi_n(t) \geq \liminf_n \varphi_n(t_n) = \varphi_\vee(1).$$

On the other hand, it is true that

$$\liminf_n \sup_{t \in [t_1, 1]} \varphi_n(t) \geq \liminf_n \sup_{t \in (t_1, 1)} \varphi_n(t) \geq \sup_{t \in (t_1, 1)} \varphi_\vee(t) = \sup_{t \in [t_1, 1)} \varphi_\vee(t)$$

by Proposition 2.26 and right-continuity of φ in t_1. Hence, it holds that

$$\liminf_n \sup_{t \in [t_1, 1]} \varphi_n(t) \geq \max\left\{\varphi_\vee(1), \sup_{t \in [t_1, 1)} \varphi_\vee(t)\right\} = \sup_{t \in [t_1, t_2]} \varphi_\vee(t).$$

This shows convergence of the suprema, precisely

$$\sup_{t \in [t_1, t_2]} \varphi_n(t) \to \sup_{t \in [t_1, t_2]} \varphi_\vee(t) = \sup_{t \in [t_1, t_2]} \varphi(t)$$

is true, where the last equality follows with Lemma 2.9, ii). Note that so far we only used the properties of the hypo-convergence of φ. Convergence of the infima, namely

$$\inf_{t \in [t_1, t_2]} \varphi_n(t) \to \inf_{t \in [t_1, t_2]} \varphi_\wedge(t) = \inf_{t \in [t_1, t_2]} \varphi(t),$$

follows similarly with the aid of epi-convergence of φ (in fact this is only a matter of signs). Theorem 2.25 finally yields $\mathrm{d}_{s,2}(\varphi_n, \varphi) \to 0$.

For the reverse implication suppose $\mathrm{d}_{s,2}(\varphi_n, \varphi) \to 0$ and choose $t_1, t_2 \in \{1\} \cup \mathrm{Disc}(\varphi)^c$. By Theorem 2.25 it holds that

$$\liminf_n \inf_{t \in [t_1, t_2]} \varphi_n(t) = \inf_{t \in [t_1, t_2]} \varphi(t) \geq \inf_{t \in [t_1, t_2]} \varphi_\wedge(t).$$

On the other hand, note that $\inf_{t\in[t_1,t_2]}\varphi_n(t) = \min\big\{\inf_{t\in(t_1,t_2)}\varphi_n(t),\, \varphi_n(t_2)\big\}$ as well as $\inf_{t\in(t_1,t_2)}\varphi_n(t) \le \varphi_n(t_2-)$ are true, which imply

$$\begin{aligned}
&\limsup_n \inf_{t\in(t_1,t_2)} \varphi_n(t) \\
&= \limsup_n \Big[\inf_{t\in[t_1,t_2]}\varphi_n(t) + \inf_{t\in(t_1,t_2)}\varphi_n(t) - \inf_{t\in[t_1,t_2]}\varphi_n(t)\Big] \\
&\le \limsup_n \inf_{t\in[t_1,t_2]}\varphi_n(t) + \limsup_n \Big[\inf_{t\in(t_1,t_2)}\varphi_n(t) - \min\Big\{\inf_{t\in(t_1,t_2)}\varphi_n(t),\, \varphi_n(t_2)\Big\}\Big] \\
&\le \limsup_n \inf_{t\in[t_1,t_2]}\varphi_n(t) + \limsup_n \Big[\varphi_n(t_2-) - \varphi_n(t_2)\Big]
\end{aligned}$$

with the aid of Lemma 2.24. The term $\big|\varphi_n(t_2-) - \varphi_n(t_2)\big|$ converges to 0 by Corollary 12.11.2, Whitt (2002), such that

$$\limsup_n \inf_{t\in(t_1,t_2)}\varphi_n(t) \le \limsup_n \inf_{t\in[t_1,t_2]}\varphi_n(t) = \inf_{t\in[t_1,t_2]}\varphi(t) = \inf_{t\in[t_1,t_2]}\varphi_\wedge(t) \le \inf_{t\in(t_1,t_2)}\varphi_\wedge(t)$$

is true by Theorem 2.25. The inequalities

$$\begin{aligned}
\limsup_n \sup_{t\in[t_1,t_2]}\varphi_n(t) &\le \sup_{t\in[t_1,t_2]}\varphi_\vee(t) \\
\liminf_n \sup_{t\in(t_1,t_2)}\varphi_n(t) &\ge \sup_{t\in(t_1,t_2)}\varphi_\vee(t)
\end{aligned}$$

follow analogously with the aid of the convergence of the suprema in Theorem 2.25 (or by looking at $-\varphi_n$ and using the inequalities shown so far). Thus, the inequalities needed for epi- and hypo-convergence in Proposition 2.26 are true for closed and open intervals respectively.

The inequalities for arbitrary open sets $G \subset [0,1]$ can be deduced from the case of open intervals by writing $G = \bigcup_{i\in\mathbb{N}}(t_1^i, t_2^i)$ with $t_1^i, t_2^i \in \{1\}\cup \mathrm{Disc}(\varphi)^c$.

For arbitrary compact K we choose a covering consisting of open intervals (t_1^i, t_2^i) with $t_1^i, t_2^i \in \{1\}\cup\mathrm{Disc}(\varphi)^c$ and $\sup_i |t_2^i - t_1^i| \le \varepsilon$. This covering has a finite sub-cover $\bigcup_{j_\varepsilon=1}^{n_\varepsilon}(t_1^{j_\varepsilon}, t_2^{j_\varepsilon})$, and taking the closure thereof yields $K \subset \bigcup_{j_\varepsilon=1}^{n_\varepsilon}[t_1^{j_\varepsilon}, t_2^{j_\varepsilon}] = I_\varepsilon$. Then, for example, it follows that

$$\liminf_n \inf_{t\in K}\varphi_n(t) \ge \inf_{t\in I_\varepsilon}\varphi(t)$$

with the aid of the assertions for closed intervals. Taking the limit $\varepsilon \searrow 0$ yields

$$\liminf_n \inf_{t\in K}\varphi_n(t) \ge \inf_{t\in K}\varphi(t). \qquad \square$$

It remains to proof Lemma 2.23, which is the M_2-equivalent of Lemma 2.13. We only sketch the ideas.

Outline for the proof of Lemma 2.23. For i) a combination of Corollary 12.11.1, ii), Theorem 12.11.6 and Corollary 12.11.6, Whitt (2002), shows that we only have to deal with the M_2-convergence of $\varphi\, h_n$. Then observe that for an arbitrary M_2-parametric representation (r, u) of h, by continuity of φ the map $t \mapsto \big(r(t), u(t)\,(\varphi \circ r)(t)\big)$ is an M_2-parametric representation for $\varphi\, h$.

Part ii) is valid as an M_2-parametric representations (r, u) of h directly converts to an M_2-parametric representation (r', u') for ${}^1\!/_h$, by keeping the time part of the parametric representation, $r' = r$, and using the reciprocal of the spatial part, $u' = {}^1\!/_u$. $\square$

Chapter 3

Asymptotics for (q_α, es_α)

In this chapter we derived the joint asymptotic distribution of empirical quantiles and Expected Shortfalls under general conditions on the distribution function of the observations. In particular, we did not assume that the distribution function is differentiable at the quantile with strictly positive derivative. Hence, the rate of convergence and the asymptotic distribution for the quantile can be non-standard, but our results showed that the Expected Shortfall remains asymptotically normal with a $\sqrt{n}$-rate, and we even gave the joint distribution in such non-standard cases. We also considered spectral risk measures with finitely-supported spectral measures and visualized our results with numerical illustrations.

3.1 Introduction

Value at Risk and Expected Shortfall are two popular measures of risk as presented in Chapter 1.

Statistical estimation of a given α-quantile, $\alpha \in (0, 1)$, is a very well-developed problem. Precise asymptotic expansions, called Bahardur expansions, for the empirical quantile have been developed if the underlying distribution function F has a density which is positive and sufficiently regular at the α-quantile (Bahadur, 1966; Kiefer, 1967). This expansion in particular implies the asymptotic normality. In this regular case an alternative quantile estimator based on a smoothed empirical distribution function has been proposed by Chen and Tang (2005) to improve finite-sample Mean-Square-Error properties. The general case in which the distribution function F is not differentiable at the α-quantile or in which its derivative vanishes was studied in Smirnov (1952) and Knight (2002). Here, non-normal limit distributions and slower rates of convergence than $\sqrt{n}$ occur.

The Expected Shortfall at level α can be estimated as the empirical average below the empirical α-quantile. Scaillet (2004) proposed to use a smoothed version of this estimator instead. Chen (2008) proved asymptotic normality of these estimators and further showed that no improvement in terms of Mean-Square-Error properties can be expected for the smoothed estimator. Further work on the asymptotic properties of the Expected Shortfall estimators are Linton and Xiao (2013) and Hill (2013) for heavy-tailed distributions and Peracchi and Tanase (2008), Taylor (2008), Cai and Wang (2008) and Kato (2012) in a nonparametric regression framework.

All these papers require that the distribution function is quite regular in its α-quantile, having a smooth and positive density as required for asymptotic normality when estimating the quantile.

In this chapter we show that this assumption is not required for the Expected Shortfall and that the simple estimator thereof remains normal under the weak assumption that the distribution function is continuous and strictly increasing at its α-quantile. We even determine the joint asymptotic distribution of the estimators for the α-quantile and Expected Shortfall in this general case. Our approach is based on the argmax-continuity theorem, stated for example in van der Vaart (1998), by using the scoring functions for the bivariate parameter (quantile, Expected Shortfall) as introduced in Example 1.8. Because of the different rates, application of the argmax-continuity theorem is not straightforward and requires substantial technical effort. This problem can be settled by applying Theorem 1.14 and Lemma 1.15.

The chapter is organised as follows. In Section 3.2 we discuss the minimum-contrast estimators defined by the chosen bivariate score for the pair (quantile, Expected Shortfall). Section 3.3 presents our results on the joint asymptotic distribution of the estimators, which is also generalized to a multivariate version considering various levels simultaneously. We further discuss asymptotic properties of estimators of spectral risk measures with finitely-supported spectral measures. Section 3.4 contains simulations in two scenarios, once for a kink in the distribution function, and once for a density with a root of order two. In Section 3.5 we summarize the results as well as indicate properties of the bootstrap, and also extensions to dependent data. Proofs of the major steps are deferred to Section 3.6, while some details are further relegated to Section 3.7.

For the rest of the chapter let $Y, Y_1, Y_2, \ldots \in \mathcal{L}_1$ denote a sequence of independent random variables distributed according to a distribution function F.

3.2 Estimating quantile and Expected Shortfall

In this section we determine the shape of the empirical estimators and compare the Expected Shortfall estimator to the empirical Tail Conditional Expectation, for which we prove that they do not vary much.

For the specific value of α under consideration we shall always impose the following.

Assumption.
For the given $\alpha \in (0,1)$, the distribution function F is continuous and strictly increasing at its α-quantile q_α. ⋆

Then F has a unique α-quantile and the empirical quantile,

$$\overline{q}_{n,\alpha} = q_\alpha(F_n) = \inf\left\{ x \Big| \sum_{i=1}^{n} \mathbb{1}\left(Y_i \leq x\right) \geq n\alpha \right\} = Y_{\lceil n\,\alpha\rceil:n},$$

is a consistent estimator for q_α (van der Vaart, 1998, Lemma 21.2). By the assumption on α it holds that $q_\alpha = q_\alpha^-$ and hence $F(q_\alpha^-) = F(q_\alpha) = \alpha$. As we saw in (1.2), in that case the Expected Shortfall coincides with the Tail Conditional Expectation meaning

$$es_\alpha = \alpha^{-1}\mathbb{E}\left[Y\,\mathbb{1}\left(Y \leq q_\alpha\right)\right] = \frac{1}{\alpha}\int_{-\infty}^{q_\alpha} y\,\mathrm{d}F(y). \tag{3.1}$$

We use the class of strictly consistent scoring functions for the bivariate parameter (q_α, es_α) as introduced in Example 1.8,

$$\begin{aligned} S(x_1, x_2; z) = {} & \left(\mathbb{1}\left(z \leq x_1\right) - \alpha\right)(x_1 - z) + G(x_2)\left(x_2 + \alpha^{-1}\left(\mathbb{1}\left(z \leq x_1\right) - \alpha\right)(x_1 - z)\right) \\ & - \mathcal{G}(x_2) - G(x_2)z, \end{aligned}$$

where $\mathcal{G}$ is a three-times continuously differentiable function, $\mathcal{G}' = G$ holds, and it is required that $G' > 0$. We choose $\mathcal{G}$ so that $\lim_{x\to-\infty} G(x) = 0$ and remember that $S(x_1, x_2; F)$ has a unique minimum in (q_α, es_α).

We consider the associated M-estimator for the parameter (q_α, es_α) defined by

$$(\widehat{q}_{n,\alpha}, \widehat{es}_{n,\alpha}) \in \underset{(x_1,x_2)\in\mathbb{R}^2}{\arg\min} \sum_{i=1}^{n} S(x_1, x_2; Y_i).$$

As the proposition below shows, this is, at least approximately, simply another way of representing standard estimators for the quantile and the Expected Shortfall. An analogous proof actually shows the representation of the Expected Shortfall in (1.2) as well, which is achieved by replacing F_n with F below.

3.1 Proposition.

The estimator $\widehat{q}_{n,\alpha}$ can be chosen equal to the empirical quantile. Further, the *estimator $\widehat{es}_{n,\alpha}$ is given by*

$$\begin{aligned}\widehat{es}_{n,\alpha} = es_\alpha(F_n) &= \arg\min_{x_2 \in \mathbb{R}} S\big(\widehat{q}_{n,\alpha}, x_2; F_n\big) \\ &= \alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le \widehat{q}_{n,\alpha})\big] + \widehat{q}_{n,\alpha}\Big(1 - \frac{1}{\alpha\, n}\sum_{i=1}^{n} \mathbb{1}\,(Y_i \le \widehat{q}_{n,\alpha})\Big)\end{aligned} \tag{3.2}$$

and we have that

$$\Big|\widehat{es}_{n,\alpha} - \alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le \widehat{q}_{n,\alpha})\big]\Big| \le \frac{\widehat{q}_{n,\alpha}}{\alpha\, n} = O_{\mathrm{P}}(n^{-1}).$$

This proposition follows from Corollary 4.3, Acerbi and Tasche (2002b); for convenience we give a proof in Section 3.6.

The empirical (lower) Tail Conditional Expectation is

$$TCE_\alpha(F_n) = \alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le \widehat{q}_{n,\alpha})\big] = \frac{1}{\alpha\, n}\sum_{i=1}^{n} \mathbb{1}\,(Y_i \le \widehat{q}_{n,\alpha})\, Y_i,$$

see the discussion after (1.2). Due to the jumps of F_n, we cannot expect an equality between $\widehat{es}_{n,\alpha}$ and $TCE_\alpha(F_n)$. But, as the proposition above shows, the estimator $\widehat{es}_{n,\alpha}$ is, up to a term of order $O_{\mathrm{P}}(n^{-1})$, equal to $TCE_\alpha(F_n)$. This is plausible as a jump of F_n in $\widehat{q}_{n,\alpha}$ vanishes asymptotically by assumption on F and the Expected Shortfall coincides with the Tail Conditional Expectation if no jump is present in the α-quantile; see (1.2) again. Thus, the asymptotic properties of $TCE_\alpha(F_n)$ will be identical to those of $\widehat{es}_{n,\alpha}$.

3.3 Joint asymptotic theory for quantile and Expected Shortfall

Here we investigate the asymptotic properties of the estimators defined before, regarding consistency and joint asymptotic distribution, which is then extended to a multivariate version. We also state asymptotic properties of M-estimators for spectral risk measures as in (1.3). We start the asymptotic analysis by providing a general consistency result.

3.2 Proposition.

Let q_n be a consistent estimator of q_α. Then the estimators

$$\alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le q_n)\big] \quad \textit{and} \quad \widetilde{es}_{n,\alpha} = \arg\min_{x_2 \in \mathbb{R}} \sum_{i=1}^{n} S(q_n, x_2; Y_i)$$

are consistent for es_α. In particular, $\big(\widehat{q}_{n,\alpha}, \widehat{es}_{n,\alpha}\big)$ is consistent.

We prove this statement in Section 3.6.

Now we turn to the joint asymptotic distribution of quantile and Expected Shortfall. One major issue is to include the case of low regularity of F at its α-quantile q_α. In particular, we do not impose the standard assumption that F has a positive derivative at q_α. In such more general settings, the possible limit distributions for the empirical quantile have been characterized in Smirnov (1952) and Knight (2002).

The standard situation of a positive derivative of F at q_α basically implies a local linearity around the point q_α. In contrast, we consider the following assumption taken from Smirnov (1952) and Knight (2002) which allows for non-linearity of F around q_α.

Assumption [A].
There exists a function $\psi_\alpha : \mathbb{R} \to \overline{\mathbb{R}}$ with

$$\lim_{t\to\infty} \psi_\alpha(t) = \infty \quad \text{and} \quad \lim_{t\to-\infty} \psi_\alpha(t) = -\infty,$$

such that for some deterministic positive sequence $(a_n)_n$ with $a_n \to \infty$ it holds that

$$\lim_{n\to\infty} \sqrt{n}[F(q_\alpha + \tfrac{t}{a_n}) - F(q_\alpha)] = \psi_\alpha(t). \qquad \star$$

In the standard case of F being differentiable in q_α with derivative $F'(q_\alpha) > 0$, Assumption [A] is fulfilled with the choice $\psi_\alpha(t) = F'(q_\alpha)t$ and $a_n = \sqrt{n}$, which gives the ordinary differential quotient.

The following proposition, which is mainly taken from Smirnov (1952), recalls the classification of the functions ψ_α which may occur in Assumption [A] and further shows that, if the empirical α-quantile is a consistent estimator for q_α, then Assumption [A] can always be satisfied with a degenerate choice for the function ψ_α.

3.3 Proposition.

i) The function ψ_α in Assumption [A] necessarily takes one of the forms

$$\psi_\alpha(t) = \begin{cases} \kappa_+ t^\beta & \text{if } t \geq 0, \\ -\infty & \text{if } t < 0, \end{cases} \qquad \psi_\alpha(t) = \begin{cases} -\kappa_-(-t)^\beta & \text{if } t \leq 0, \\ \infty & \text{if } t > 0, \end{cases}$$

$$\psi_\alpha(t) = \begin{cases} -\kappa_-(-t)^\beta & \text{if } t \leq 0, \\ \kappa_+ t^\beta & \text{if } t > 0, \end{cases} \qquad \psi_\alpha(t) = \begin{cases} -\infty & \text{if } t < -c_1, \\ 0 & \text{if } -c_1 \leq t \leq c_2, \\ \infty & \text{if } t > c_2, \end{cases}$$

where $\kappa_+, \kappa_-, \beta > 0$ and $c_1, c_2 \geq 0$. Moreover, except for the last case with $c_1 = c_2 = 0$, which means $\psi_\alpha(t) = \infty \cdot \operatorname{sign}(t)$ with $\infty \cdot 0 = 0$, the sequence $(a_n)_n$ is uniquely determined up to asymptotic equivalence.

ii) If the empirical α-quantile is consistent for q_α, then there exists a sequence (a_n) for which Assumption [A] is satisfied for the limit function $\psi_\alpha(t) = \infty \cdot \operatorname{sign}(t)$.

Here, sequences of positive numbers $(a_n)_n$ and $(b_n)_n$ are *asymptotically equivalent* if $\frac{a_n}{b_n} \to 1$ as $n \to \infty$. Part ii) of the proposition implies that Assumption [A] imposes no additional general restrictions if F is strictly increasing and continuous at its α-quantile. The proof can be found in Section 3.6.

Having a distribution function F, which admits a right- and left-sided derivative $F^+(q_\alpha), F^-(q_\alpha) > 0$ in q_α, Assumption [A] is valid using $\kappa_- = F^-(q_\alpha)$, $\kappa_+ = F^+(q_\alpha)$ and $\beta = 1$ in the third case of the former proposition. In this situation, the parameter β expresses a sort of local linearity around q_α.

More general, β indicates the polynomial behaviour in a neighbourhood of q_α as the following example shows.

3.4 Example.
Assume that there exists an $\varepsilon > 0$ and functions κ_+, κ_- which are continuous in q_α with $\kappa_+(q_\alpha)$, $\kappa_-(q_\alpha) \neq 0$ and fulfil

$$\begin{aligned} F(x) - \alpha &= (x - q_\alpha)^{r+1}\,\kappa_+(x) \quad \text{for } x \in [q_\alpha, q_\alpha + \varepsilon) \quad \text{and} \\ F(x) - \alpha &= (q_\alpha - x)^{l+1}\,\kappa_-(x) \quad \text{for } x \in (q_\alpha - \varepsilon, q_\alpha] \end{aligned}$$

for some $r, l \in (-1, \infty)$. For example, if F has a density with a root of order $k \in \mathbb{N}_0$ in its α-quantile, these assertions are met; see Sections 3.4.1 and 3.4.2.

Since we assume strict monotonicity of F in q_α, we must have $\kappa_+(x) > 0$ for $x \in (q_\alpha, q_\alpha + \varepsilon)$ and hence $\kappa_+(q_\alpha) > 0$ as well (it is $\neq 0$ by assumption). Similarly, $\kappa_-(x) < 0$ is valid for $x \in (q_\alpha - \varepsilon, q_\alpha]$. Then, setting $a_n^\vee = n^{1/(2(r+1))}$, for $t > 0$ we have that $q_\alpha + \frac{t}{a_n^\vee} \in [q_\alpha, q_\alpha + \varepsilon)$ for n big enough, hence

$$\sqrt{n}\Big(F\big(q_\alpha + \tfrac{t}{a_n^\vee}\big) - F\big(q_\alpha\big)\Big) = \sqrt{n}\,\kappa_+\big(q_\alpha + \tfrac{t}{a_n^\vee}\big)\,\frac{t^{r+1}}{\sqrt{n}} \longrightarrow \kappa_+(q_\alpha)t^{r+1} > 0 \text{ as } n \to \infty$$

is true. Similarly, for $t < 0$ and $a_n^\wedge = n^{1/(2(l+1))}$ we have that

$$\sqrt{n}\Big(F\big(q_\alpha + \tfrac{t}{a_n^\wedge}\big) - F\big(q_\alpha\big)\Big) \longrightarrow \kappa_-(q_\alpha)(-t)^{l+1} < 0.$$

Now, if $r = l$, we can choose $\beta = r + 1$, $a_n = n^{1/(2\beta)}$ and

$$\psi_\alpha(t) = \begin{cases} \kappa_-\big(q_\alpha\big)\,(-t)^\beta & \text{if } t \leq 0, \\ \kappa_+\big(q_\alpha\big)\,t^\beta & \text{if } t > 0. \end{cases}$$

Then the sequence a_n together with the function ψ_α fulfil Assumption [A].

If $r > l$, choosing $a_n = n^{1/(2\beta)}$ and $\beta = r+1$ again, we have for $t < 0$ and n big enough that
$$\sqrt{n}\Big(F\big(q_\alpha + \tfrac{t}{a_n}\big) - F(q_\alpha)\Big) = n^{(r-l)/(2\beta)}\,\kappa_-\big(q_\alpha + \tfrac{t}{a_n}\big)\,(-t)^{l+1} \longrightarrow -\infty.$$
Thus, Assumption [A] is satisfied in this case for $a_n = n^{1/(2\beta)}$ and
$$\psi_\alpha(t) = \begin{cases} -\infty & \text{if } t < 0, \\ \kappa_+(q_\alpha)t^\beta & \text{if } t \geq 0. \end{cases}$$
The case $l > r$ is treated similarly with $\beta = l+1$. ◇

The next assumption ensures the existence of a limit variance for the estimator $\widehat{es}_{n,\alpha}$.

Assumption [B].
It holds that $\mathbb{E}\big[\mathbb{1}\,(Y \leq 0)\; Y^2\big] < \infty$.

⋆

Now we can state our main result for this chapter, whose proof is relegated to Section 3.6.

3.5 Theorem.

Under Assumptions [A] and [B], we have that
$$\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha), \sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)\big) \xrightarrow{\mathcal{L}} \big(\psi_\alpha^{\leftrightarrow}(W_1), W_2\big),$$
where (W_1, W_2) *are jointly normally distributed,*
$$(W_1, W_2) \sim \mathcal{N}(0, \Sigma), \quad \Sigma = \begin{pmatrix} \alpha(1-\alpha) & (1-\alpha)(q_\alpha - es_\alpha) \\ (1-\alpha)(q_\alpha - es_\alpha) & \frac{1}{\alpha^2}\mathrm{Var}\big(\mathbb{1}\,(Y \leq q_\alpha)\,(q_\alpha - Y)\big) \end{pmatrix},$$
and
$$\psi_\alpha^{\leftrightarrow}(x) = \begin{cases} \inf\{t \leq 0 \,|\, \psi_\alpha(t) \geq x\} & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ \sup\{t \geq 0 \,|\, \psi_\alpha(t) \leq x\} & \text{if } x > 0. \end{cases} \tag{3.3}$$

The theorem implies that the marginal asymptotic distribution of the estimator $\widehat{es}_{n,\alpha}$ is not affected by low regularity of the distribution function F at q_α, although the rate of convergence and asymptotic distribution of $\widehat{q}_{n,\alpha}$ become non-standard. This is not unsurprising for the following reason. For a known value q_α of the α-quantile we could consider the oracle estimator
$$\frac{1}{\alpha\, n}\sum_{i=1}^{n} \mathbb{1}\,(Y_i \leq q_\alpha)\, Y_i,$$

which has asymptotic variance $\frac{1}{\alpha^2}\mathrm{Var}\big(\mathbb{1}\,(Y \le q_\alpha)\,Y\big)$. Now it holds that

$$\tfrac{1}{\alpha^2}\mathrm{Var}\big(\mathbb{1}\,(Y \le q_\alpha)\,(q_\alpha - Y)\big) - \tfrac{1}{\alpha^2}\mathrm{Var}\big(\mathbb{1}\,(Y \le q_\alpha)\,Y\big) = \frac{1-\alpha}{\alpha}\,q_\alpha\,(q_\alpha - 2es_\alpha).$$

If $q_\alpha < 0$, which is plausible in applications since we consider the lower tail Expected Shortfall, then since $es_\alpha \le q_\alpha$ the difference is negative so that estimating the quantile actually may reduce the asymptotic variance of the Expected Shortfall. This effect persists even if it is quite hard – as in situations with low regularity of F at q_α – to estimate the quantile.

3.6 Remark.
Chen and Tang (2005) proposed a smoothed estimator of the quantile and showed that higher-order correction of the Mean-Square-Error is possible for an appropriate choice of the bandwidth. Scaillet (2004) proposed a smoothed estimator of the Expected Shortfall, but the asymptotic analysis in Chen (2008) showed that no asymptotic improvement can be expected. Thus, Chen (2008) recommends the use of the simple empirical Expected Shortfall. What is more, the favourable analysis of Chen and Tang (2005) for the smoothed estimator of the quantile depends on regularity of F and q_α, roughly a twice-continuously differentiable density. We shall investigate the behaviour of this smoothed estimator of the Expected Shortfall in our less regular situations in the numerical illustrations in Section 3.4. ◇

In Example 3.4 we saw that β indicates how “smooth” F is in a neighbourhood of the considered quantile. As we see now, β is closely linked to the rate of convergence of $\widehat{q}_{n,\alpha}$.

3.7 Example (Example 3.4 continued).
Consider the situation of Example 3.4, and additionally assume that Assumption [B] is satisfied. If $r = l$, Theorem 3.5 applies with $a_n = n^{1/(2\beta)}$, $\beta = r + 1$, and

$$\psi_\alpha^{\leftrightarrow}(u) = \begin{cases} -\big(\frac{u}{\kappa_-(q_\alpha)}\big)^{1/\beta} & \text{if } u < 0, \\ 0 & \text{if } u = 0, \\ \big(\frac{u}{\kappa_+(q_\alpha)}\big)^{1/\beta} & \text{if } u > 0. \end{cases}$$

For $r > l$ Theorem 3.5 still applies with $a_n = n^{1/(2\beta)}$, where β is as before, but in the formula for $\psi_\alpha^{\leftrightarrow}(u)$ above we have to replace the case $u < 0$ with $\psi_\alpha^{\leftrightarrow}(u) = 0$. We can deal with $r < l$, $\beta = l + 1$ analogously. ◇

Next let us extend Theorem 3.5 to a multivariate version. For given $k \in \mathbb{N}$ choose distinct $\alpha_s \in (0,1)$, $s \in \{1, \dots, k\}$, and assume as before that F is strictly monotone and continuous at each quantile q_{α_s}.

Assumption [A^k].
For each $s \in \{1, \dots, k\}$ and corresponding α_s and q_{α_s}, Assumption [A] is satisfied with associated sequence $(a_{s,n})_n$ and function $\psi_{\alpha_s}(t)$. ⋆

Under this assumption we are able to generalize Theorem 3.5 as follows.

3.8 Theorem.

Let Assumptions [A^k] and [B] hold. Then the vector

$$\Big(a_{1,n}\big(\widehat{q}_{n,\alpha_1}-q_{\alpha_1}\big),\sqrt{n}\,\big(\widehat{es}_{n,\alpha_1}-es_{\alpha_1}\big),\ldots,a_{k,n}\big(\widehat{q}_{n,\alpha_k}-q_{\alpha_k}\big),\sqrt{n}\,\big(\widehat{es}_{n,\alpha_k}-es_{\alpha_k}\big)\Big)$$

converges weakly to $(z_{1,1},z_{1,2},\ldots,z_{k,1},z_{k,2})$. Here, for $s=1,\ldots,k$, $z_{s,1}=\psi^{\leftrightarrow}_{\alpha_s}(W_{s,1})$ and $z_{s,2}=W_{s,2}$, where $\psi^{\leftrightarrow}_{\alpha_s}$ is as in (3.3) and the vector $(W_{1,1},W_{1,2},\ldots,W_{k,1},W_{k,2})$ is distributed according to $\mathcal{N}(0,\Sigma)$ with Σ for $s,t\in\{1,\ldots,k\}$ determined by

$$\begin{aligned}
\mathrm{Cov}\big(W_{s,1},W_{t,1}\big)&=\alpha_s\wedge\alpha_t-\alpha_s\alpha_t,\\
\mathrm{Cov}\big(W_{s,2},W_{t,2}\big)&=\frac{\alpha_s\wedge\alpha_t}{\alpha_s\alpha_t}\big(q_{\alpha_s}q_{\alpha_t}-(q_{\alpha_s}+q_{\alpha_t})es_{\alpha_s\wedge\alpha_t}\big)\\
&\quad+\frac{1}{\alpha_s\alpha_t}\mathbb{E}\big[\mathbb{1}\,(Y\le q_{\alpha_s\wedge\alpha_t})\,Y^2\big]+\big(es_{\alpha_s}-q_{\alpha_s}\big)\big(es_{\alpha_t}-q_{\alpha_t}\big),\\
\mathrm{Cov}\big(W_{s,2},W_{t,1}\big)&=\frac{\alpha_s\wedge\alpha_t}{\alpha_t}\big(q_{\alpha_t}-es_{\alpha_s\wedge\alpha_t}\big)-\alpha_s\big(q_{\alpha_t}-es_{\alpha_t}\big).
\end{aligned}$$

The extension of the proof of Theorem 3.5 to the multivariate case in Theorem 3.8 is relegated to Section 3.7.

As an application of the above theorem consider the estimation of spectral risk measures with spectral measure having finite support. Therefore let m be a probability measure on $[0,1]$, finitely supported in $(0,1)$, and κ_m the associated spectral risk measure as in (1.3),

$$\kappa_m=\sum_{s=1}^{k}p_s\,es_{\alpha_s}\quad\text{if}\quad m=\sum_{s=1}^{k}p_s\delta_{\alpha_s}.$$

In Example 1.8 we got to know the strictly consistent scoring functions for κ_m given by

$$\begin{aligned}
S_{sp}(x_1,\ldots,x_k,x_{k+1};z)=\sum_{s=1}^{k}\Bigg(&\Big(1+\frac{p_s}{\alpha_s}G(x_{k+1})\Big)\big(\mathbb{1}\,(z\le x_s)-\alpha\big)\,(x_s-z)\\
&+p_s\big(G(x_{k+1})(x_{k+1}-z)-\mathcal{G}(x_{k+1})\big)\Bigg).
\end{aligned}$$

We then have the following result for the corresponding M-estimator

$$\big(\widehat{q}_{n,\alpha_1},\ldots,\widehat{q}_{n,\alpha_k},\widehat{\kappa}_{m,n}\big)\in\underset{x_1,\ldots,x_{k+1}\in\mathbb{R}}{\arg\min}\;S_{sp}(x_1,\ldots,x_{k+1};F_n).$$

3.9 Theorem.

We have that

$$\widehat{\kappa}_{m,n} = \sum_{s=1}^{k} p_s\, \widehat{es}_{n,\alpha_s}. \tag{3.4}$$

Consequently, under Assumptions [A^k] and [B] it follows that

$$\sqrt{n}\left(\widehat{\kappa}_{m,n} - \kappa_m\right) \xrightarrow{\mathcal{L}} \sum_{s=1}^{k} p_s\, W_{s,2},$$

where the $W_{s,2}$ are as in Theorem 3.8.

3.4 Numerical illustrations

Here we complete the theoretical results obtained before with a short simulation study. Therefore we choose two distribution functions fulfilling Assumptions [A] and [B] needed in Theorem 3.5 and simulate observations thereof using the statistic software `R`. From the observations we then calculate the empirical distribution functions of the quantile and Expected Shortfall estimates and compare them to their theoretical limits.

3.4.1 Distribution function with kink in the α-quantile

We let F be given by

$$F(x) = \frac{1}{5}(x+1)\,\mathbb{1}\left(x \in (-1,0]\right) + \left(\frac{1}{5} + \frac{8}{5}\,x\right)\,\mathbb{1}\left(x \in (0,1/2]\right) + \mathbb{1}\left(x \in (1/2,\infty)\right).$$

The distribution function possesses a kink in 0, which is the $\frac{1}{5}$-quantile. So let us fix $\alpha = 1/5$ and estimate $q_{1/5} = 0$ and the Expected Shortfall $es_{1/5} = -1/2$. We observe that the left- and right-sided derivatives of F in $q_{1/5}$ are given by $F^-(0) = 1/5$ and $F^+(0) = 8/5$.

A Taylor expansion shows that Example 3.4 applies with $r = l = 0$, $a_n = \sqrt{n}$ and

$$\psi_{1/5}(t) = t\left(F^-(0)\,\mathbb{1}\left(t \le 0\right) + F^+(0)\,\mathbb{1}\left(t \ge 0\right)\right),$$

so that

$$\psi_{1/5}^{\leftrightarrow}(t) = t\,\left(F^-(0)^{-1}\,\mathbb{1}\left(t \le 0\right) + F^+(0)^{-1}\,\mathbb{1}\left(t \ge 0\right)\right)$$

is valid. It follows with Theorem 3.5 that

$$\sqrt{n}\begin{pmatrix} \widehat{q}_{n,1/5} - 0 \\ \widehat{es}_{n,1/5} + 1/2 \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} W_1\left[F^-(0)^{-1}\mathbb{1}\left(W_1 < 0\right) + F^+(0)^{-1}\,\mathbb{1}\left(W_1 > 0\right)\right] \\ W_2 \end{pmatrix},$$

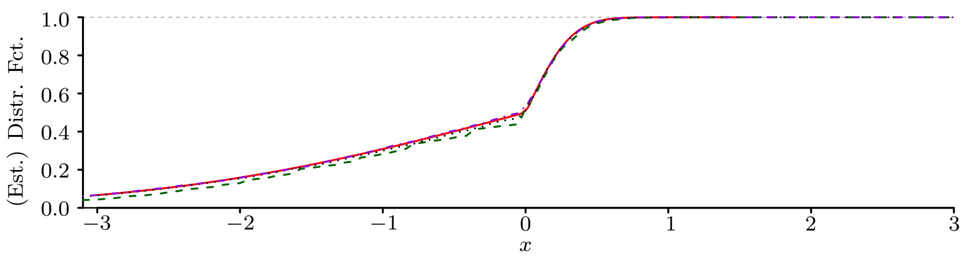
where

$$(W_1, W_2) \sim \mathcal{N}\left(0, \begin{pmatrix} 4/25 & 2/5 \\ 2/5 & 5/3 \end{pmatrix}\right).$$

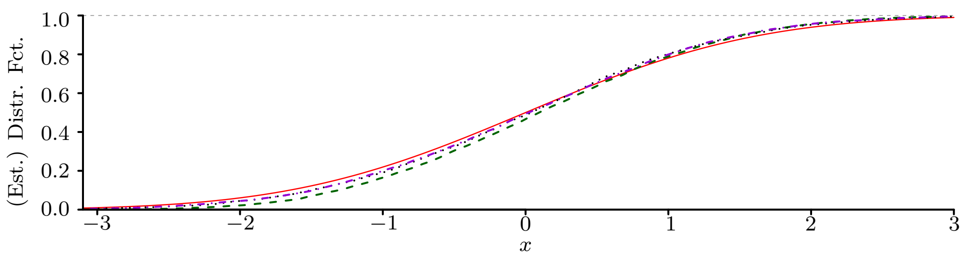
The limit distribution function of $\sqrt{n}\,\widehat{q}_{n,1/5}$ is calculated as

$$z \mapsto \Phi_{0,4/25}\Big(z\,(F^-(0)\,\mathbb{1}\,(t \le 0) + F^+(0)\,\mathbb{1}\,(t \ge 0))\Big);$$

the one of $\sqrt{n}\,(\widehat{es}_{n,1/5} + 1/2)$ is $\Phi_{0,5/3}$.



(a) Estimated quantile distribution function.



(b) Estimated Expected Shortfall distribution function.

Figure 3.1: Picture (a) shows the limit (red solid) and the estimated distribution functions of $\sqrt{n}\,\widehat{q}_{n,1/5}$ for the distribution function of Example 3.4.1 for $n = 10^2$ (green dashed), $n = 10^3$ (black dotted), and $n = 10^4$ (purple dot-dashed). Picture (b) accordingly shows the limit and the estimated distribution functions of $\sqrt{n}\big(\widehat{es}_{n,1/5} + 1/2\big)$ using the same colour code.

The M-estimators $\widehat{q}_{n,1/5}$ and $\widehat{es}_{n,1/5}$ were computed for simulated samples of sizes $n \in \{10^2, 10^3, 10^4, 5 \cdot 10^4, 10^5, 10^6\}$, each for $5 \cdot 10^3$ iterations. This procedure was then repeated $2 \cdot 10^2$ times in order to obtain more reliable characteristics of the estimated distributions; see Tables 3.1, 3.2, 3.3 and 3.4. We depict the asymptotic and one of the estimated distribution functions of $\sqrt{n}\,\widehat{q}_{n,1/5}$ and $\sqrt{n}\,(\widehat{es}_{n,1/5} + 1/2)$ for samples of sizes $n \in \{10^2, 10^3, 10^4\}$ (Figure 3.1). The approximation is reasonable in both cases also for small sample sizes.

From the same data we in addition computed the smoothed quantile estimator $\widetilde{q}_{h_n,n,1/5}$ and the estimator $\widetilde{es}_{h_n,n,1/5}$ for the Expected Shortfall as proposed in Chen and Tang

(2005) and Chen (2008) respectively. Here we used fixed bandwidths h_n chosen as the median normal reference bandwidth of additional training samples.

We observe that the limit distribution of $\sqrt{n}\,\widehat{q}_{n,1/5}$ does not have mean 0 (Table 3.1), while the mean of $\sqrt{n}\,\widetilde{q}_{h_n,n,1/5}$ seems to diverge (Table 3.1). Smoothing the Expected Shortfall also appears to introduce a small bias, which seems to decay slowly (Table 3.2).

Table 3.1: Mean and standard deviation of the (smoothed) rescaled and centred estimator $\widehat{q}_{n,1/5}$. The values were calculated depending on $5 \cdot 10^3$ estimates coming from samples of size n (choosing a decreasing bandwidth $h_n \in \{0.082, 0.052, 0.0333, 0.024, 0.0201, 0.0132\}$ for the smoothed version $\widehat{q}_{h_n,n,1/5}$), and were last averaged over $2 \cdot 10^2$ repetitions of that scheme; the bracketed numbers show the resulting standard deviations.

Size	Mean		Standard deviation	
n	$\sqrt{n}\,\widehat{q}_{n,1/5}$	$\sqrt{n}\,\widetilde{q}_{h_n,n,1/5}$	$\sqrt{n}\,\widehat{q}_{n,1/5}$	$\sqrt{n}\,\widetilde{q}_{h_n,n,1/5}$
10^2	−0.587 (0.015)	−0.906 (0.013)	1.115 (0.014)	0.940 (0.013)
10^3	−0.668 (0.017)	−1.497 (0.013)	1.205 (0.017)	0.899 (0.013)
10^4	−0.690 (0.017)	−2.799 (0.012)	1.232 (0.017)	0.833 (0.010)
$5{\cdot}10^4$	−0.695 (0.017)	−4.430 (0.011)	1.234 (0.016)	0.815 (0.008)
10^5	−0.702 (0.018)	−5.233 (0.012)	1.232 (0.017)	0.811 (0.008)
10^6	−0.739 (0.017)	−10.801 (0.012)	1.213 (0.016)	0.813 (0.008)
True	−0.698	−0.698	1.353	1.353

Last, we report averaged quantiles of the estimated distribution functions of $\sqrt{n}\,\widehat{q}_{n,1/5}$ and $\sqrt{n}\,(\widehat{es}_{n,1/5} + \frac{1}{2})$ (Tables 3.3 and 3.4). In case of the quantile estimator, the true values are approximated well, whereas the estimator for the Expected Shortfall mostly gives higher values for quantiles below the 0.5-quantile and higher values for the above. This can also be seen in Figure 3.1 where the estimated distribution functions tend to lie below the true one for $x < 0$ and above else.

3.4.2 Density with root of order 2

Let us fix $\alpha = 1/2$ and

$$F(x) = \mathbb{1}\,(x \in [0,2])\,\frac{((x-1)^3+1)}{2} + \mathbb{1}\,(x \in (2,\infty))\,.$$

Then $F(1) = 1/2$, so that $q_{1/2} = 1$ and $es_{1/2} = 1/4$. Example 3.4 applies with $r = l = 2$, $\varepsilon = 1$ and $\kappa_+(x) = -\kappa_-(x) = 1/2$, hence $a_n = n^{1/6}$ and $\psi_{1/2}(t) = t^3/2$ together satisfy

Table 3.2: Mean and standard deviation of the (smoothed) rescaled and centred estimator $\widehat{es}_{n,1/5}$. The values were obtained from on $5 \cdot 10^3$ estimates coming from samples of size n (with decreasing bandwidth $h_n \in \{0.082, 0.052, 0.0333, 0.024, 0.0201, 0.0132\}$ for the smoothed version $\widehat{es}_{h_n,n,1/5}$), and were last averaged over $2 \cdot 10^2$ repetitions of the described scheme; the numbers in brackets show the calculated standard deviations.

Size	Mean		Standard deviation	
n	$\sqrt{n}\left(\widehat{es}_{n,1/5} + \frac{1}{2}\right)$	$\sqrt{n}\left(\widetilde{es}_{h_n,n,1/5} + \frac{1}{2}\right)$	$\sqrt{n}\left(\widehat{es}_{n,1/5} + \frac{1}{2}\right)$	$\sqrt{n}\left(\widetilde{es}_{h_n,n,1/5} + \frac{1}{2}\right)$
10^2	0.105 (0.015)	0.185 (0.016)	1.083 (0.010)	1.112 (0.010)
10^3	0.036 (0.016)	0.136 (0.017)	1.153 (0.011)	1.172 (0.011)
10^4	0.011 (0.017)	0.144 (0.017)	1.18 (0.012)	1.189 (0.012)
$5{\cdot}10^4$	0.008 (0.017)	0.165 (0.017)	1.184 (0.012)	1.189 (0.012)
10^5	0.002 (0.017)	0.157 (0.017)	1.183 (0.012)	1.186 (0.012)
10^6	0.001 (0.017)	0.215 (0.017)	1.188 (0.011)	1.189 (0.011)
True	0.000	0.000	1.291	1.291

Table 3.3: Quantiles of the estimated distribution function of $\sqrt{n}\,\widehat{q}_{n,1/5}$, based on $5 \cdot 10^3$ samples of sizes n and averaged over $2 \cdot 10^2$ repetitions of the procedure. The small numbers denote the resulting standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	−3.98 (0.070)	−2.904 (0.048)	−2.303 (0.059)	−1.208 (0.019)	0.000 (0.000)	0.172 (0.006)	0.341 (0.008)	0.444 (0.010)	0.642 (0.015)
10^3	−4.444 (0.096)	−3.174 (0.058)	−2.486 (0.046)	−1.316 (0.037)	0.000 (0.001)	0.172 (0.005)	0.328 (0.006)	0.423 (0.007)	0.604 (0.014)
10^4	−4.595 (0.101)	−3.257 (0.056)	−2.544 (0.046)	−1.337 (0.038)	−0.008 (0.018)	0.170 (0.005)	0.323 (0.006)	0.414 (0.008)	0.587 (0.013)
$5{\cdot}10^4$	−4.618 (0.094)	−3.269 (0.057)	−2.549 (0.047)	−1.335 (0.037)	−0.010 (0.020)	0.169 (0.005)	0.322 (0.006)	0.412 (0.007)	0.583 (0.013)
10^5	−4.621 (0.107)	−3.266 (0.058)	−2.549 (0.047)	−1.348 (0.041)	−0.015 (0.022)	0.166 (0.005)	0.321 (0.006)	0.411 (0.007)	0.580 (0.013)
10^6	−4.627 (0.105)	−3.276 (0.059)	−2.558 (0.047)	−1.341 (0.039)	−0.013 (0.020)	0.000 (0.000)	0.314 (0.005)	0.377 (0.004)	0.580 (0.014)
True	−4.653	−3.290	−2.563	−1.349	0.000	0.169	0.320	0.411	0.582

Assumption [A]. Assumption [B] is fulfilled as well with $4\,\mathrm{Var}\big(\mathbb{1}\,(Y \leq 1)\,(1-Y)\big) = {}^{51}\!/_{80}$. The map $\psi_{1/2}$ is invertible with $\psi^{\leftrightarrow}_{1/2}(y) = \psi^{\mathrm{Inv}}_{1/2}(y) = (2y)^{1/3}$ and thus, using Theorem 3.5

Table 3.4: Quantiles of the estimated distribution function of $\sqrt{n}\,(\widehat{es}_{n,1/5} + 1/2)$, based on $5 \cdot 10^3$ samples of sizes n and averaged over $2 \cdot 10^2$ repetitions of the scheme; the numbers in brackets indicate the resulting standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	−2.202 (0.040)	−1.626 (0.028)	−1.288 (0.023)	−0.669 (0.020)	0.078 (0.021)	0.850 (0.022)	1.533 (0.026)	1.931 (0.034)	2.640 (0.054)
10^3	−2.571 (0.053)	−1.845 (0.034)	−1.442 (0.028)	−0.756 (0.021)	0.027 (0.019)	0.817 (0.023)	1.526 (0.028)	1.944 (0.035)	2.729 (0.059)
10^4	−2.710 (0.059)	−1.924 (0.033)	−1.499 (0.028)	−0.789 (0.024)	0.008 (0.022)	0.808 (0.023)	1.526 (0.029)	1.955 (0.037)	2.754 (0.059)
$5{\cdot}10^4$	−2.734 (0.062)	−1.938 (0.040)	−1.510 (0.029)	−0.793 (0.023)	0.008 (0.022)	0.808 (0.023)	1.525 (0.030)	1.954 (0.035)	2.764 (0.066)
10^5	−2.740 (0.060)	−1.939 (0.036)	−1.512 (0.029)	−0.798 (0.023)	0.000 (0.021)	0.800 (0.023)	1.518 (0.028)	1.949 (0.036)	2.753 (0.061)
10^6	−2.749 (0.060)	−1.950 (0.033)	−1.520 (0.027)	−0.800 (0.022)	0.000 (0.021)	0.803 (0.024)	1.526 (0.028)	1.957 (0.033)	2.762 (0.062)
True	−3.003	−2.123	−1.654	−0.871	0.000	0.871	1.654	2.123	3.003

we obtain

$$\begin{pmatrix} n^{1/6}\,(\widehat{q}_{n,1/2} - 1) \\ \sqrt{n}\,(\widehat{es}_{n,1/2} - 1/4) \end{pmatrix} \xrightarrow{\mathcal{L}} \begin{pmatrix} \psi^{\mathrm{Inv}}_{1/2}(W_1) \\ W_2 \end{pmatrix} = \begin{pmatrix} (2\,W_1)^{1/3} \\ W_2 \end{pmatrix}.$$

Here, W_1 and W_2 are distributed according to

$$(W_1, W_2) \sim \mathcal{N}(0, \Sigma), \qquad \Sigma = \begin{pmatrix} 1/4 & 3/8 \\ 3/8 & 51/80 \end{pmatrix}.$$

The distribution function of $\psi^{\mathrm{Inv}}_{1/2}(W_1)$ is $\Phi_{0,1/4}\big(\frac{1}{2}t^3\big)$; the one of W_2 is $\Phi_{0,51/80}$. Additionally, the joint density of the pair $\big(\psi^{\mathrm{Inv}}_{1/2}(W_1),\, W_2\big)$ is given by

$$f_{\psi^{\mathrm{Inv}}_{1/2}(W_1),W_2}(t,v) = \frac{3\,t^2}{4\,\pi\sqrt{\det\Sigma}}\exp\left(-\frac{1}{8}(t^3\ v)\,\Sigma^{-1}\begin{pmatrix} t^3 \\ v \end{pmatrix}\right),$$

using the transformation formula for integrals.

The considered estimators were computed for $n \in \{10^2, 10^3, 10^4, 5\cdot 10^4, 10^5, 10^6\}$, again each for $5\cdot 10^3$ iterations. This was then repeated $2\cdot 10^2$ times to obtain the characteristics summarized in Tables 3.5, 3.6 and 3.7.

We visually compare the true distribution functions with exemplary estimated distribution functions of $n^{1/6}\,(\widehat{q}_{n,1/2} - 1)$ and $\sqrt{n}\,(\widehat{es}_{n,1/2} - 1/4)$, choosing $n \in \{10^2, 10^3, 10^4\}$ for the quantile estimator (Figure 3.2 (a)) as well as for the Expected Shortfall estimator (Figure 3.2 (b)). The estimated distribution functions in the quantile case fluctuate less around the true distribution than in the Expected Shortfall case. For the chosen estimators, the estimated distribution functions of $\sqrt{n}\,(\widehat{es}_{n,1/2} - 1/4)$ always lie below the true curve. This seems to be quite persistent throughout the iterations, as the (averaged) estimated quantiles of $\sqrt{n}\,(\widehat{es}_{n,1/2} - 1/4)$ always lie above the quantiles of the limit W_2 (Table 3.7).

Additionally, we report the empirical bias and standard deviation of the rescaled and centred estimators $\widehat{q}_{n,1/2}$ and $\widehat{es}_{n,1/2}$ averaged over $2 \cdot 10^2$ repetitions (Table 3.5). The values are quite stable and seem to converge.

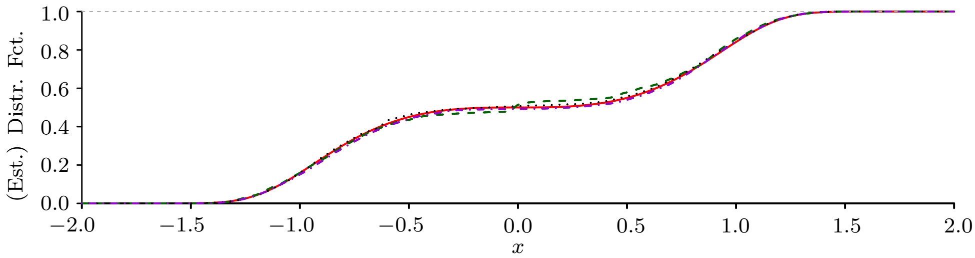
The visual impression of convergence is further supported by the quantiles of the estimated distribution functions, which are close to the quantiles of the asymptotic distribution in the quantile case (Tables 3.6 and 3.7).

Overall, the asymptotic approximation is reasonable for the quantile already for moderate sample sizes but the Expected Shortfall requires quite large sample sizes for the asymptotic approximation to become valid.

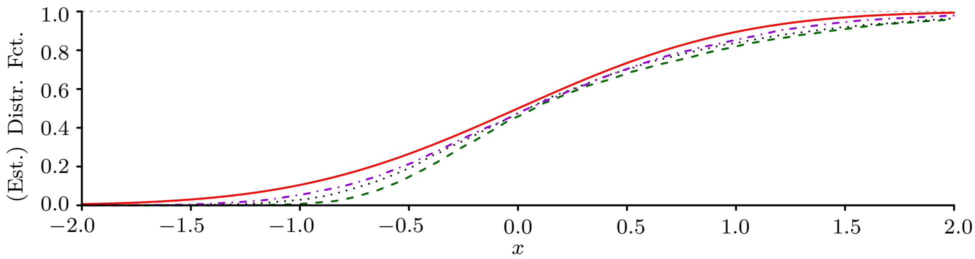
In Figure 3.3 we used $n = 10^6$ and increased the number of iterations to $5 \cdot 10^4$ in order to nonparametrically estimate the joint density function of the bivariate random variable $\big(n^{1/6}(\widehat{q}_{n,1/2} - 1), \sqrt{n}(\widehat{es}_{n,1/2} - 1/4)\big)$ using the `R`-package `ks`. This density estimate then is compared to the asymptotic density $f_{\psi^{\mathrm{Inv}}_{1/2}(W_1),W_2}$ calculated above. The shape of the true density is captured well.

Table 3.5: Mean and standard deviation of the rescaled and centred estimators $\widehat{q}_{n,1/2}$ and $\widehat{es}_{n,1/2}$. The values were calculated from $5 \cdot 10^3$ estimates coming from samples of size n and were last averaged over $2 \cdot 10^2$ repetitions of the scheme. The small numbers in brackets show the resulting standard deviations.

Size	Mean		Standard deviation	
n	$n^{1/6}\big(\widehat{q}_{n,1/2} - 1\big)$	$\sqrt{n}\big(\widehat{es}_{n,1/2} - \frac{1}{4}\big)$	$n^{1/6}\big(\widehat{q}_{n,1/2} - 1\big)$	$\sqrt{n}\big(\widehat{es}_{n,1/2} - \frac{1}{4}\big)$
10^2	−0.008 (0.013)	0.288 (0.012)	0.883 (0.004)	0.837 (0.010)
10^3	−0.001 (0.013)	0.199 (0.011)	0.893 (0.003)	0.820 (0.010)
10^4	0.000 (0.013)	0.134 (0.011)	0.896 (0.003)	0.810 (0.009)
$5{\cdot}10^4$	0.000 (0.012)	0.103 (0.011)	0.896 (0.003)	0.804 (0.008)
10^5	0.000 (0.012)	0.092 (0.011)	0.896 (0.003)	0.803 (0.009)
10^6	0.000 (0.013)	0.063 (0.011)	0.896 (0.003)	0.800 (0.009)
True	0.000	0.000	0.896	0.798



(a) Estimated quantile distribution function.



(b) Estimated Expected Shortfall distribution function.

Figure 3.2: Picture (a) shows the limit (red solid) and the estimated distribution function of $n^{1/6}\big(\widehat{q}_{n,1/2} - 1\big)$ for Section 3.4.2 with $n = 10^2$ (green dashed), $n = 10^3$ (black dotted), and $n = 10^4$ (purple dot-dashed), while picture (b) presents the limit and the estimated distribution function of $\sqrt{n}\big(\widehat{es}_{n,1/2} - 1/4\big)$ with the same colour code.

3.5 Conclusions and discussion

We showed that the assumption of having a positive density at the α-quantile, required for the quantile estimate to be asymptotically normal at $\sqrt{n}$-rate, is *not* required for asymptotic normality of the Expected Shortfall.

The asymptotic variance of the Expected Shortfall can be estimated by forming a sample-counterpart expression. Alternatively, one may use the bootstrap. For the quantile in non-standard situations Knight (1998) shows that the simple n-out-of-n bootstrap is not consistent but sub-sampling works. For the marginal asymptotic distribution of the Expected Shortfall, however, additional simulations indicate that the n-out-of-n bootstrap is consistent, even without regularity of the density at the quantile.

In this chapter we only considered independent identically distributed data. Quantile and Expected Shortfall estimation is often applied to financial time series, and therefore extensions of the results to dependent data would be useful. These should be possible but the details, in particular general M-estimation theory based on dependent data by using the argmax-continuity theorem, still need to be developed.

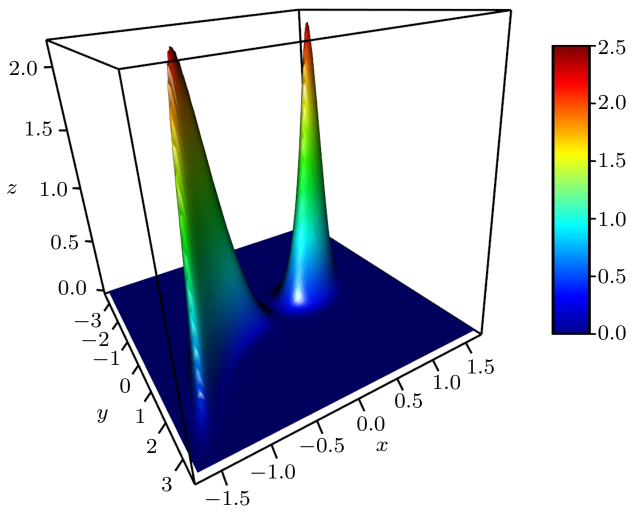
Table 3.6: Quantiles of the estimated distribution function of $n^{1/6}\,(\widehat{q}_{n,1/2} - 1)$, calculated from $5 \cdot 10^3$ samples of sizes n. Last, the resulting empirical quantiles were averaged over $2 \cdot 10^2$ repetitions of the procedure, where the numbers in brackets are the resulting standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	−1.306 (0.010)	−1.183 (0.008)	−1.087 (0.008)	−0.876 (0.008)	−0.001 (0.012)	0.876 (0.010)	1.079 (0.007)	1.173 (0.006)	1.315 (0.01)
10^3	−1.322 (0.009)	−1.178 (0.007)	−1.084 (0.007)	−0.874 (0.009)	−0.068 (0.135)	0.877 (0.008)	1.086 (0.007)	1.180 (0.007)	1.324 (0.009)
10^4	−1.324 (0.009)	−1.181 (0.007)	−1.087 (0.007)	−0.878 (0.008)	0.000 (0.212)	0.877 (0.008)	1.086 (0.007)	1.181 (0.007)	1.325 (0.010)
$5{\cdot}10^4$	−1.323 (0.010)	−1.179 (0.007)	−1.086 (0.007)	−0.877 (0.008)	0.000 (0.214)	0.876 (0.008)	1.086 (0.007)	1.181 (0.007)	1.323 (0.010)
10^5	−1.323 (0.010)	−1.179 (0.007)	−1.085 (0.006)	−0.877 (0.008)	−0.003 (0.224)	0.877 (0.009)	1.086 (0.007)	1.180 (0.008)	1.324 (0.011)
10^6	−1.326 (0.009)	−1.180 (0.007)	−1.086 (0.007)	−0.876 (0.008)	−0.001 (0.240)	0.877 (0.009)	1.086 (0.007)	1.181 (0.007)	1.324 (0.009)
True	−1.325	−1.18	−1.086	−0.877	0.000	0.877	1.086	1.180	1.325

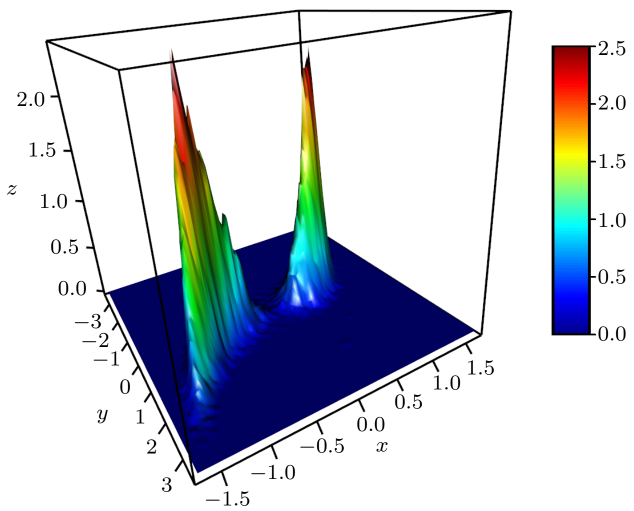
Table 3.7: Quantiles of the estimated distribution function of $\sqrt{n}\,(\widehat{es}_{n,1/2} - 1/4)$, based on $5{\cdot}10^3$ samples of sizes n. Finally, the empirical quantiles were averaged over $2{\cdot}10^2$ repetitions of the scheme; the numbers in brackets indicate the resulting standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	−0.948 (0.017)	−0.723 (0.011)	−0.587 (0.009)	−0.325 (0.009)	0.077 (0.014)	0.748 (0.023)	1.478 (0.030)	1.932 (0.035)	2.809 (0.069)
10^3	−1.213 (0.026)	−0.898 (0.015)	−0.719 (0.013)	−0.393 (0.011)	0.051 (0.014)	0.680 (0.019)	1.340 (0.026)	1.749 (0.036)	2.536 (0.061)
10^4	−1.413 (0.027)	−1.029 (0.017)	−0.816 (0.014)	−0.442 (0.012)	0.033 (0.014)	0.633 (0.018)	1.236 (0.025)	1.609 (0.032)	2.323 (0.058)
$5{\cdot}10^4$	−1.507 (0.031)	−1.091 (0.019)	−0.863 (0.015)	−0.465 (0.012)	0.025 (0.014)	0.609 (0.018)	1.186 (0.023)	1.537 (0.028)	2.206 (0.053)
10^5	−1.546 (0.033)	−1.114 (0.018)	−0.878 (0.016)	−0.472 (0.013)	0.022 (0.014)	0.603 (0.018)	1.167 (0.025)	1.515 (0.032)	2.169 (0.055)
10^6	−1.647 (0.035)	−1.179 (0.021)	−0.924 (0.017)	−0.493 (0.013)	0.016 (0.014)	0.583 (0.018)	1.122 (0.022)	1.449 (0.029)	2.064 (0.046)
True	−1.857	−1.313	−1.023	−0.539	0.000	0.539	1.023	1.313	1.857

Finally, the analysis of the Expected Shortfall as a process in the level α would be of some interest, in particular to study general spectral risk measures when not assuming a finitely-supported spectral measure.



(a) Joint density of limit distribution.

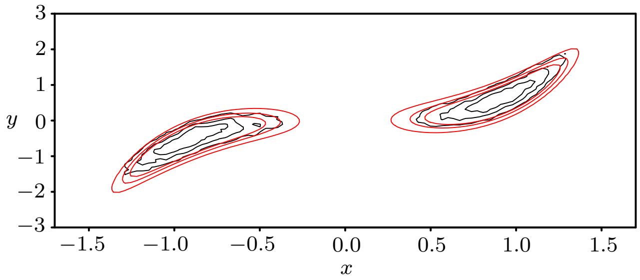


(b) Estimated joint density.

3.6 Proofs

Here we give the proofs and important intermediate results needed to obtain the assertions in Sections 3.2 and 3.3. We start by proving the propositions stated until here and proceed by presenting the cornerstones needed to conclude Theorem 3.5.

Some technical details are shifted to Section 3.7 in order to focus on the important steps required.



(c) Contour plots.

Figure 3.3: The images show (a) the limit joint density function and (b) the estimated joint density function of $\big(n^{1/6}(\widehat{q}_{n,1/2} - 1), \sqrt{n}(\widehat{es}_{n,1/2} - 1/4)\big)$. Image (c) shows the contour lines (75%, 50%, 25% from outer to inner lines) of the theoretical (red) and the estimated (black) density in the above example.

3.6.1 Proofs of Propositions 3.1, 3.2 and 3.3

To start with, observe that we may write the score for (q_α, es_α) in (3.2) equivalently as

$$S(x_1, x_2; z) = (1 + \alpha^{-1} G(x_2)) \left(\mathbb{1}\left(z \leq x_1\right) - \alpha\right)(x_1 - z) + G(x_2)\left(x_2 - z\right) - \mathcal{G}(x_2). \tag{3.5}$$

The course for proving Proposition 3.1 is similar to the proof of Theorem 1.9. We cannot directly apply this here as F_n does not have unique quantiles.

Proof of Proposition 3.1. Define the functions

$$\rho_\alpha(x_1; z) = \left(\mathbb{1}\left(z \leq x_1\right) - \alpha\right)(x_1 - z), \qquad g(x_2) = \left(1 + \alpha^{-1} G(x_2)\right),$$
$$h(x_2; z) = G(x_2)(x_2 - z) - \mathcal{G}(x_2)$$

so that

$$S(x_1, x_2; z) = g(x_2)\rho_\alpha(x_1; z) + h(x_2; z),$$

see (3.5), hence

$$(\widehat{q}_{n,\alpha}, \widehat{es}_{n,\alpha}) \in \underset{(x_1,x_2)\in\mathbb{R}^2}{\arg\min} \Big\{ g(x_2)\rho_\alpha(x_1; F_n) + h(x_2; F_n) \Big\}$$

holds. The minimal value equals

$$\min_{x_2\in\mathbb{R}} \Big\{ g(x_2)\Big(\min_{x_1\in\mathbb{R}} \rho_\alpha(x_1; F_n)\Big) + h(x_2; F_n) \Big\}, \tag{3.6}$$

so that the minimizer in the first coordinate does not depend on the choice of x_2. As ρ_α is a score for the quantile, namely the check function from Example 1.7, it follows that

$$\widehat{q}_{n,\alpha} \in \arg\min_{x_1 \in \mathbb{R}} \rho_\alpha(x_1; F_n) = [Y_{\lceil n\,\alpha \rceil : n}, Y_{\lceil n\,\alpha \rceil + 1 : n}],$$

which includes the empirical quantile $\overline{q}_{n,\alpha}$. This is the first part of the assertion.

From (3.6), $\widehat{es}_{n,\alpha}$ minimizes

$$x_2 \mapsto S\big(\widehat{q}_{n,\alpha}, x_2; F_n\big) = g(x_2)\rho_\alpha(\widehat{q}_{n,\alpha}; F_n) + h(x_2; F_n).$$

The partial derivatives of the functions g and h are given by

$$\partial_{x_2} g(x_2) = \alpha^{-1} G'(x_2) \quad \text{and} \quad \partial_{x_2} h(x_2; z) = G'(x_2)(x_2 - z),$$

respectively, thus

$$\begin{aligned}
\partial_{x_2} \sum_{i=1}^{n} S\big(\widehat{q}_{n,\alpha}, x_2; Y_i\big) &= G'(x_2) \sum_{i=1}^{n} \Big[\alpha^{-1} \rho_\alpha(\widehat{q}_{n,\alpha}; Y_i) + x_2 - Y_i\Big] \\
&= G'(x_2) \sum_{i=1}^{n} \Big[x_2 - \widehat{q}_{n,\alpha} + \alpha^{-1} \mathbb{1}\left(Y_i \le \widehat{q}_{n,\alpha}\right)\left(\widehat{q}_{n,\alpha} - Y_i\right)\Big]
\end{aligned}$$

is valid. As $G'(x_2) > 0$ by assumption, setting the above derivative equal to zero is equivalent to

$$0 = n\, x_2 - n\, \widehat{q}_{n,\alpha} + \alpha^{-1} \sum_{i=1}^{n} \mathbb{1}\left(Y_i \le \widehat{q}_{n,\alpha}\right)\left(\widehat{q}_{n,\alpha} - Y_i\right).$$

By multiplying this with n^{-1} and reorganising the resulting equation the second claim follows.

For the final estimate, we observe that by the above calculations the equality

$$\Big|\widehat{es}_{n,\alpha} - \alpha^{-1} \mathbb{E}_n\big[Y\, \mathbb{1}\left(Y \le \widehat{q}_{n,\alpha}\right)\big]\Big| = \alpha^{-1} \widehat{q}_{n,\alpha} \Big|\alpha - \mathbb{E}_n\big[\mathbb{1}\left(Y \le \widehat{q}_{n,\alpha}\right)\big]\Big|$$

holds. So it remains to discuss $\big|\alpha - \mathbb{E}_n\big[\mathbb{1}\left(Y \le \widehat{q}_{n,\alpha}\right)\big]\big|$. As $\widehat{q}_{n,\alpha} \in [Y_{\lceil n\,\alpha \rceil : n}, Y_{\lceil n\,\alpha \rceil + 1 : n}]$ is true, we know that $\mathbb{E}_n\big[\mathbb{1}\left(Y \le \widehat{q}_{n,\alpha}\right)\big] \le (\lceil n\,\alpha \rceil + 1)/n$ and thus we obtain

$$\Big|\alpha - \mathbb{E}_n[\mathbb{1}\left(Y \le \widehat{q}_{n,\alpha}\right)]\Big| \le \frac{1}{n}\big(\lceil n\,\alpha \rceil + 1 - n\,\alpha\big) \le \frac{2}{n},$$

which implies the assertion. □

Next, we prove the consistency of $(\widehat{q}_{n,\alpha}, \widehat{es}_{n,\alpha})$, which was formulated slightly more general in Proposition 3.2.

Proof of Proposition 3.2. By the law of large numbers and the definition (3.1) of es_α

$$\Big|\alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le q_\alpha)\,\big] - es_\alpha\Big| = o_{\mathrm{P}}(1)$$

is valid. This implies that

$$\Big|\alpha^{-1}\mathbb{E}_n[Y\,\mathbb{1}\,(Y \le q_n)] - es_\alpha\Big| = \alpha^{-1}\Big|\mathbb{E}_n\big[Y\big(\mathbb{1}\,(Y \le q_n) - \mathbb{1}\,(Y \le q_\alpha)\,\big)\big]\Big| + o_{\mathrm{P}}(1),$$

and it remains to show the convergence

$$\Big|\mathbb{E}_n\big[Y\big(\mathbb{1}\,(Y \le q_n) - \mathbb{1}\,(Y \le q_\alpha)\,\big)\big]\Big| \xrightarrow{\mathrm{P}} 0. \tag{3.7}$$

Recall that as $\alpha \in (0,1)$ the quantile q_α is finite, $|q_\alpha| + 1 \le c < \infty$. Now, let $\eta > 0$ and choose $1 \ge \delta > 0$ such that $F(q_\alpha + \delta) - F(q_\alpha - \delta) \le \frac{\alpha\,\eta}{2\,c}$ – this is possible since F is continuous in q_α. On the set $\{|q_n - q_\alpha| \le \delta\}$ the integral in (3.7) is smaller than (or equal to)

$$\begin{aligned}\max\{|q_\alpha - \delta|, |q_\alpha + \delta|\}\,&\mathbb{E}_n\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big]\\ &\le c\,\mathbb{E}_n\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big].\end{aligned}$$

Then, note that by the strong law of large numbers $\mathbb{E}_n\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big]$ converges in probability to $\mathbb{E}[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)]$. Thus, it follows that

$$\begin{aligned}&\mathrm{P}\Big(\Big|\mathbb{E}_n\big[Y\big(\mathbb{1}\,(Y \le q_n) - \mathbb{1}\,(Y \le q_\alpha)\,\big)\big]\Big| \ge \eta\Big)\\ &\le \mathrm{P}\Big(|q_n - q_\alpha| \ge \delta\Big) + \mathrm{P}\Big(\mathbb{E}_n\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big] \ge \tfrac{\eta}{c}\Big)\\ &\le \mathrm{P}\Big(|q_n - q_\alpha| \ge \delta\Big) + \mathrm{P}\Big(\big|(\mathbb{E}_n - \mathbb{E})\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big]\big| \ge \tfrac{\eta}{2c}\Big).\end{aligned}$$

The last two probabilities can be made small by choosing n big enough since both $|q_n - q_\alpha| = o_{\mathrm{P}}(1)$ and $\big|(\mathbb{E}_n - \mathbb{E})\big[\mathbb{1}\,(Y \le q_\alpha + \delta) - \mathbb{1}\,(Y \le q_\alpha - \delta)\,\big]\big| = o_{\mathrm{P}}(1)$ do hold. This proves (3.7).

For the statement concerning $\widetilde{es}_{n,\alpha}$, as in the proof of Proposition 3.1 we obtain the generalization of (3.2),

$$\widetilde{es}_{n,\alpha} = \alpha^{-1}q_n\big(\alpha - \mathbb{E}_n\big[\mathbb{1}\,(Y \le q_n)\,\big]\big) + \alpha^{-1}\mathbb{E}_n\big[Y\,\mathbb{1}\,(Y \le q_n)\,\big].$$

Since from the first part of the proof, the last term above converges to es_α in probability, it remains to show $|\alpha^{-1}q_n(\alpha - \mathbb{E}_n[\mathbb{1}\,(Y \le q_n)\,])| = o_{\mathrm{P}}(1)$. For this it suffices to show $|\alpha - \mathbb{E}_n[\mathbb{1}\,(Y \le q_n)\,]| = o_{\mathrm{P}}(1)$, as $\alpha^{-1}\,q_n$ is tight by assumption. The argument for this remaining convergence is the same as for (3.7), what concludes the proof of the proposition. □

Last for this subsection we deal with the properties concerning Assumption [A].

Proof of Proposition 3.3. Ad i). The stated classification of ψ_α is shown in Smirnov (1952, § 4). Uniqueness of (a_n) up to asymptotic equivalence follows from the convergence of types theorem and the distributional convergence of $a_n(\widehat{q}_{n,\alpha} - q_\alpha)$ to a non-degenerate limit distribution under Assumption [A]; see Knight (2002) or the proof of Theorem 3.5.

Ad ii). If $(Y_{\lceil n\,\alpha\rceil:n} - q_\alpha) = o_{\mathrm{P}}(1)$, then one can find a sequence $a_n \to \infty$ for which $a_n(Y_{\lceil n\,\alpha\rceil:n} - q_\alpha) = o_{\mathrm{P}}(1)$ is still true. By Theorem 4, Smirnov (1952), this holds if and only if

$$\frac{F(q_\alpha + {}^{t}\!/\!{}_{a_n}) - \lambda_{n,\alpha}}{\tau_{n,\alpha}} \to u(t). \tag{3.8}$$

Here, $u : \mathbb{R} \to \overline{\mathbb{R}}$ is a non-decreasing function uniquely determined by

$$\mathbb{1}\,(t \in [0, \infty)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u(t)} \exp\Big(-\tfrac{x^2}{2}\Big)\,\mathrm{d}x,$$

further

$$\lambda_{n,\alpha} = \frac{\lceil n\,\alpha\rceil}{n+1}, \quad \tau_{n,\alpha} = \sqrt{\frac{\lambda_{n,\alpha}\iota_{n,\alpha}}{n+1}} \quad \text{and} \quad \iota_{n,\alpha} = \frac{n - \lceil n\,\alpha\rceil + 1}{n+1}.$$

With these definitions note that

$$\lambda_{n,\alpha} \to \alpha \quad \text{and} \quad \iota_{n,\alpha} \to 1 - \alpha$$

hold. Thus, the convergence in (3.8) is equivalent to

$$\frac{\sqrt{n+1}\big(F(q_\alpha + {}^{t}\!/\!{}_{a_n}) - \alpha\big)}{\sqrt{\alpha(1-\alpha)}} \to u(t)$$

which then yields the convergence stated in Assumption [A] with a_n as chosen above and $\psi_\alpha(t) = \sqrt{\alpha(1-\alpha)}\,u(t)$. □

3.6.2 Proof of Theorem 3.5

Now, we give the main parts for proving Theorem 3.5; most proofs are relegated to the end of this section in order to better understand the scheme. The proof is divided into the following five steps.

Step 1 Considering increments of the scoring function.

Step 2 Determining the rate of convergence of $\widehat{es}_{n,\alpha}$.

Step 3 Weak convergence of the process to be minimized.

Step 4 Approximation of $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$ with other minimizers.

Step 5 Conclusion with the aid of the argmax-continuity theorem.

Steps 1–4 are proven by utilizing Theorem 1.14 and Lemma 1.15 and enable Step 5.

Step 1 Increments of the scoring function.

The shape of the increments of the score S is comprised in the following lemma.

3.10 Lemma.

i) We have that

$$\begin{aligned}
&S(x_1, x_2 + y_2; z) - S(x_1, x_2; z) \\
&= \big(G(x_2 + y_2) - G(x_2)\big)\big(x_2 - x_1 + \tfrac{1}{\alpha}\mathbb{1}\,(z \le x_1)\,(x_1 - z)\big) \\
&\quad + \int_0^{y_2} G'(x_2 + s) s \,\mathrm{d}s \qquad (3.9) \\
&= \big(G(x_2 + y_2) - G(x_2)\big)\big(x_2 - x_1 + \tfrac{1}{\alpha}\mathbb{1}\,(z \le x_1)\,(x_1 - z)\big) \\
&\quad + \tfrac{1}{2}G'(x_2 + y_2)y_2^2 - \tfrac{1}{2}\int_0^{y_2} G''(x_2 + s)s^2 \,\mathrm{d}s. \qquad (3.10)
\end{aligned}$$

ii) Setting $\rho_\alpha(x_1; z) = \big(\mathbb{1}\,(z \le x_1) - \alpha\big)(x_1 - z)$ it holds that

$$\begin{aligned}
&\rho_\alpha(x_1 + y_1; z) - \rho_\alpha(x_1; z) \\
&\quad = y_1\big(\mathbb{1}\,(z \le x_1) - \alpha\big) + \int_0^{y_1} \big(\mathbb{1}\,(z \le x_1 + s) - \mathbb{1}\,(z \le x_1)\big)\,\mathrm{d}s. \qquad (3.11)
\end{aligned}$$

iii) Generally, the equality

$$\begin{aligned}
&S(x_1 + y_1, x_2 + y_2; z) - S(x_1, x_2; z) \\
&= \big(1 + \tfrac{1}{\alpha}G(x_2 + y_2)\big)\Big[y_1\big(\mathbb{1}\,(z \le x_1) - \alpha\big) + \int_0^{y_1} \mathbb{1}\,(z \le x_1 + s) - \mathbb{1}\,(z \le x_1)\,\mathrm{d}s\Big] \\
&\quad + \big(G(x_2 + y_2) - G(x_2)\big)\big(x_2 - x_1 + \tfrac{1}{\alpha}\mathbb{1}\,(z \le x_1)\,(x_1 - z)\big) \\
&\quad + \tfrac{1}{2}G'(x_2 + y_2)y_2^2 - \tfrac{1}{2}\int_0^{y_2} G''(x_2 + s)s^2 \,\mathrm{d}s \qquad (3.12)
\end{aligned}$$

is valid.

The proof of Lemma 3.10 is given in Section 3.7. We will use the definition of ρ_α in ii), which is the strictly consistent scoring function for the quantile from Example 1.7, through the following.

Step 2 Determining the rate of convergence of $\widehat{es}_{n,\alpha}$.

The following lemma is proved by checking the assumptions of Theorem 1.14, which is further sketched below and rigorously carried out in Section 3.7. The statement is slightly more general than needed, as we do not focus on $\widehat{q}_{n,\alpha}$ alone.

3.11 Lemma.

Assume q_n to be a consistent estimator of q_α and Assumption [B] to hold. Then the sequence $\sqrt{n}(\widetilde{es}_{n,\alpha} - es_\alpha)$ is tight, where $\widetilde{es}_{n,\alpha}$ is the minimizer of the function

$$x_2 \mapsto \sum_{i=1}^n S(q_n, x_2; Y_i) = n\mathbb{E}_n[S(q_n, x_2; Y)].$$

In particular, if Assumptions [A] and [B] hold, then $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$ is a tight sequence.

Step 3 Convergence of processes to be minimized.

Using (3.12) with $x_1 = q_\alpha$, $y_1 = \frac{u_1}{a_n}$, $x_2 = es_\alpha$ and $y_2 = \frac{u_2}{\sqrt{n}}$, where $a_n > 0$, we may write

$$\begin{aligned}&\sum_{i=1}^n S(q_\alpha + {}^{u_1}/_{a_n}, es_\alpha + {}^{u_2}/_{\sqrt{n}}; Y_i) - S(q_\alpha, es_\alpha; Y_i)\\ &= (1 + \alpha^{-1} G(es_\alpha + {}^{u_2}/_{\sqrt{n}}))V_n(u_1) + U_n(u_2),\end{aligned}$$

where V_n and U_n are defined by

$$\begin{aligned}V_n(u_1) &= \frac{u_1}{a_n}\sum_{i=1}^n (\mathbb{1}(Y_i \le q_\alpha) - \alpha) + \frac{1}{a_n}\int_0^{u_1} \Big(\sum_{i=1}^n \mathbb{1}(Y_i \le q_\alpha + {}^{t}/_{a_n}) - \mathbb{1}(Y_i \le q_\alpha)\Big)\,\mathrm{d}t\\ &= \sum_{i=1}^n \rho_\alpha(q_\alpha + {}^{u_1}/_{a_n}; Y_i) - \rho_\alpha(q_\alpha; Y_i) \qquad (3.13)\end{aligned}$$

and

$$\begin{aligned}U_n(u_2) &= \sqrt{n}(G(es_\alpha + {}^{u_2}/_{\sqrt{n}}) - G(es_\alpha))\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(es_\alpha - q_\alpha + \alpha^{-1}\mathbb{1}(Y_i \le q_\alpha)(q_\alpha - Y_i)\right)\\ &\quad + \frac{u_2^2}{2}G'(es_\alpha + {}^{u_2}/_{\sqrt{n}}) - \frac{1}{2\sqrt{n}}\int_0^{u_2} G''(es_\alpha + {}^{t}/_{\sqrt{n}})t^2\,\mathrm{d}t\\ &= \sum_{i=1}^n S(q_\alpha, es_\alpha + {}^{u_2}/_{\sqrt{n}}; Y_i) - S(q_\alpha, es_\alpha; Y_i). \qquad (3.14)\end{aligned}$$

Here we used (3.11) and (3.10) and made a substitution in the integrals.

Under Assumptions [A] and [B] the processes U_n and the rescaled processes $\frac{a_n}{\sqrt{n}} V_n$ converge in distribution as stated next. We prove this at the end of this section.

3.12 Lemma.

If Assumption [A] holds, then

$$\frac{a_n}{\sqrt{n}}\, V_n(u_1) \rightsquigarrow u_1 W_1 + \int_0^{u_1} \psi_\alpha(t)\,\mathrm{d}t =: V(u_1) \tag{3.15}$$

in $\big(\ell^\infty(K_1), \|\cdot\|_{K_1}\big)$ for every compact set $K_1 \subset \mathbb{R}$, where $W_1 \sim \mathcal{N}(0, \alpha(1-\alpha))$.
If Assumption [B] holds, we have the convergence

$$U_n(u_2) \rightsquigarrow G'(es_\alpha)\,\big(\,u_2\, W_2 + {}^{u_2^2}\!/\!_2\big) =: U(u_2) \tag{3.16}$$

in $\big(\ell^\infty(K_2), \|\cdot\|_{K_2}\big)$ for every compact set $K_2 \subset \mathbb{R}$. Here, the random variable W_2 is distributed according to $\mathcal{N}\big(0, \alpha^{-2}\operatorname{Var}_F[\mathbb{1}\,(Y \le q_\alpha)\,(q_\alpha - Y)]\big)$.
Moreover, if both Assumptions [A] and [B] hold, we have that

$$\big(\tfrac{a_n}{\sqrt{n}} V_n, U_n\big) \rightsquigarrow (V, U) \tag{3.17}$$

in $\big(\ell^\infty(K), \|\cdot\|_K\big)$ for every compact $K \subset \mathbb{R}^2$, where (W_1, W_2) in the definition of (V, U) are jointly normally distributed with mean 0, covariance $(1-\alpha)\,(q_\alpha - es_\alpha)$ and variances as before.

Step 4 Approximation of $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$ with other minimizers.

To conclude Theorem 3.5, we look at the minimizers of U_n, which have a similar asymptotic behaviour as $\widehat{es}_{n,\alpha}$.

3.13 Lemma.

The processes (U_n) and U in Lemma 3.12 have unique minimizers $(u_{2,n})$ and u_2^0, respectively, and $u_{2,n} \xrightarrow{\mathcal{L}} u_2^0$. Moreover, we have that

$$\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha) = u_{2,n} + o_{\mathrm{P}}(1). \tag{3.18}$$

The proof is deferred to the end of the section.

Step 5 Application of the argmax-continuity theorem.

From Lemma 3.13, we have that

$$\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha), \sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)\big) = \big(a_n(\widehat{q}_{n,\alpha} - q_\alpha), u_{2,n}\big) + o_{\mathrm{P}}(1).$$

Now, $\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha), u_{2,n}\big)$ is by construction a sequence of minimizers of the processes $\big(V_n(u_1) + U_n(u_2)\big)$, but, since the variables are separated, also of the processes

$$Z_n(u_1, u_2) = \big(1 + \alpha^{-1} G(es_\alpha)\big)\, \frac{a_n}{\sqrt{n}}\, V_n(u_1) + U_n(u_2),$$

which, by Lemma 3.12, (3.17) and the continuous mapping theorem, converge in the space $\big(\ell^\infty(K), \|\cdot\|_K\big)$ for any compact $K \subset \mathbb{R}^2$ to the process

$$Z(u_1, u_2) = \big(1 + \alpha^{-1} G(es_\alpha)\big)\, V(u_1) + U(u_2).$$

To conclude$\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha), u_{2,n}\big) \xrightarrow{\mathcal{L}} (z_1, z_2)$, the limit being the minimizer of Z, we apply the argmax-continuity theorem, for example Corollary 5.58, van der Vaart (1998), and need to check the remaining assumptions for this.

The process U has almost surely continuous sample paths, which is immediate from the shape given in (3.16), and a unique minimizer as stated in Lemma 3.13; further $u_{2,n}$ is a tight sequence by Lemma 3.13.

The process V also has a unique minimum almost surely. Indeed, the form of the functions $\psi_\alpha(t)$ as given in Proposition 3.3, in particular the assertion that $\kappa_+, \kappa_-, \beta > 0$, as well as the form of V in (3.15) imply that $\lim_{u_1 \to \pm\infty} V(u_1) = \infty$ and that for the closed interval for which $|V(u_1)| < \infty$ the derivative has at most one zero; if it has no zero the minimizer is on the boundary of this interval. Observe in addition, that V almost surely has continuous paths on that closed interval (with left-continuity on the right endpoint and right-continuity on the left endpoint). Moreover, in the proof of Lemma 3.13 we will see that $a_n(\widehat{q}_{n,\alpha} - q_\alpha)$ is a tight sequence. An application of the argmax-continuity theorem concludes the proof of Theorem 3.5. ☐

3.6.3 Proofs of intermediate results

Here we prove the results needed in steps 2–4; the tedious calculation for step 1 is relegated to Section 3.7. We start by presenting an outline for the proof of Lemma 3.11, which proceeds by checking the assumptions of Theorem 1.14. Detailed calculations are deferred to Section 3.7 in order to focus on the main ideas.

Proof of Lemma 3.11 (Outline). We use Theorem 1.14 with $\alpha = 2$, $\beta = 1$, d_0, d_1 the Euclidean distance in $\mathbb{R}$ and the criterion function $m_{\eta,\vartheta}(z) = S(\eta, \vartheta; z)$. Consistency of $\widehat{es}_{n,\alpha}$ for es_α has been taken care of in Theorem 3.2.

Concerning (1.7) we need to prove that

$$\inf_{|\eta - q_\alpha| \le \varepsilon} \inf_{\delta_0 \ge |\vartheta - es_\alpha| \ge \delta} \mathbb{E}\left[S(\eta, \vartheta; Y) - S(\eta, es_\alpha; Y)\right] \ge C\, \delta^2 \tag{3.19}$$

is true for all $0 < \varepsilon \leq \varepsilon_0$, $0 < \delta \leq \delta_0$ and some $C, \delta_0, \varepsilon_0 > 0$. To this end, using (3.9) in Lemma 3.10, we get by convexity and strict consistency (for q_α) of the function $\eta \mapsto \rho_\alpha(\eta; F)$ that

$$\inf_{|\eta - q_\alpha| \leq \varepsilon} \inf_{\delta_0 \geq |\varepsilon_2| \geq \delta} \mathbb{E}\left[S(\eta, es_\alpha + \varepsilon_2; Y) - S(\eta, es_\alpha; Y)\right] \geq \frac{\delta^2}{2}\Big(G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3}\Big)$$

for

$$C_1 = \sup_{x_2 \in [es_\alpha - \delta_0, es_\alpha + \delta_0]} |G''(x_2)| < \infty.$$

Since $\lim_{\delta \to 0} G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3} = G'(es_\alpha) > 0$ holds by assumption on G' we can find $C_2 > 0$ and $\delta_0 > 0$ with $G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3} \geq C_2$ for every $\delta \leq \delta_0$. This proves (3.19) with constant $C = \frac{C_2}{2}$.

For (1.8) we require

$$\mathbb{E}\Bigg[\sup_{\substack{|\eta - q_\alpha| \leq \varepsilon \\ |\vartheta - es_\alpha| \leq \delta}} \Big|\mathbb{G}_n\big[S(\eta, \vartheta; Y) - S(\eta, es_\alpha; Y)\big]\Big|\Bigg] \leq C\,\delta \tag{3.20}$$

for all $0 < \delta \leq \delta_0$ and some $\delta_0, C > 0$. Proving (3.20) reduces to showing that

$$\mathbb{E}\left[\sup_{|\eta - q_\alpha| \leq \varepsilon} \big|\mathbb{G}_n[\mathbb{1}\,(Y \leq \eta)\,(\eta - Y)]\big|\right] \leq C$$

for some constant C not depending on δ, which may be accomplished by using a maximal inequality involving the bracketing integral. $\square$

Next we prove the weak convergence of V_n and U_n using the theory of van der Vaart (1998).

Proof of Lemma 3.12. Let Assumption [A] be true. In fact, the convergence (3.15) was shown in Knight (2002), but for convenience we give a (different) proof here. First, we observe that

$$\big|\rho_\alpha(x_1; z) - \rho_\alpha(x_1'; z)\big| \leq (1 + \alpha)|x_1 - x_1'|. \tag{3.21}$$

Indeed, if $x_1 \leq x_1' < z$, then $\big|\rho_\alpha(x_1; z) - \rho_\alpha(x_1'; z)\big| = \alpha(x_1' - x_1)$ is satisfied. Else, if $x_1 < z \leq x_1'$, then

$$\begin{aligned} \big|\rho_\alpha(x_1; z) - \rho_\alpha(x_1'; z)\big| &= |-\alpha\, x_1 - x_1' + \alpha\, x_1' + z| \leq \alpha(x_1' - x_1) + (x_1' - z) \\ &\leq \alpha(x_1' - x_1) + (x_1' - x_1) = (1 + \alpha)(x_1' - x_1) \end{aligned}$$

is valid. In the last case where $z \leq x_1 \leq x_1'$ it holds that $\big|\rho_\alpha(x_1; z) - \rho_\alpha(x_1'; z)\big| = (1 - \alpha)(x_1' - x_1)$. All three cases together prove (3.21).

Using the Lipschitz continuity (3.21), from Lemma 19.31 in van der Vaart (1998) we obtain that

$$\mathbb{G}_n\Big[a_n\big(\rho_\alpha(q_\alpha + {}^{u_1}\!/_{a_n}; Y) - \rho_\alpha(q_\alpha; Y)\big) - u_1\big(\mathbb{1}\,(Y \le q_\alpha) - \alpha\big)\Big] = o_{\mathrm{P}}(1)$$

holds in $\big(\ell^\infty(K_1), \|\cdot\|_{K_1}\big)$ for any compact $K_1 \subset \mathbb{R}$ with $q_\alpha \in K_1$. Therefore, from the definition of $V_n(u_1)$ in (3.13),

$$\begin{aligned}\frac{a_n}{\sqrt{n}}V_n(u_1) &= \sqrt{n}\,\mathbb{E}_n\big[a_n\big(\rho_\alpha(q_\alpha + {}^{u_1}\!/_{a_n}; Y) - \rho_\alpha(q_\alpha; Y)\big)\big] \\ &= \frac{u_1}{\sqrt{n}}\sum_{i=1}^n \big(\mathbb{1}\,(Y_i \le q_\alpha) - \alpha\big) + \int_0^{u_1} \sqrt{n}\big(F(q_\alpha + {}^{s}\!/_{a_n}) - F(q_\alpha)\big)\,\mathrm{d}s + o_{\mathrm{P}}(1).\end{aligned} \tag{3.22}$$

The first term converges by the central limit theorem to $u_1 W_1$ for W_1 as stated. For the second, note that under Assumption [A] we also have

$$\lim_{n\to\infty}\int_0^{u_1} \sqrt{n}\big(F(q_\alpha + {}^{s}\!/_{a_n}) - F(q_\alpha)\big)\,\mathrm{d}s = \int_0^{u_1} \psi_\alpha(s)\,\mathrm{d}s.$$

Indeed, using the monotonicity of $\sqrt{n}\big(F(q_\alpha + {}^{s}\!/_{a_n}) - F(q_\alpha)\big)$, this follows from the dominated convergence theorem if $|\psi_\alpha(t)| < \infty$, and, if $|\psi_\alpha(t)| = \infty$, the fact that $\{\psi_\alpha = \infty\}$ and $\{\psi_\alpha = -\infty\}$ are open intervals (see Proposition 3.3, i)).

Now let Assumption [B] be true. Below we show that for every compact set K_2 with $es_\alpha \in K_2$ there exists a function $L(z)$ such that for every $x_2, x_2' \in K_2$,

$$\big|S(q_\alpha, x_2'; z) - S(q_\alpha, x_2; z)\big| \le L(z)\big|x_2' - x_2\big|, \tag{3.23}$$

where $L(z)$ fulfils $\mathbb{E}[L(Y)^2] < \infty$. To deduce (3.15) we then can apply Lemma 19.31 of van der Vaart (1998). We thus obtain that

$$\mathbb{G}_n\Big[\sqrt{n}\big(S(q_\alpha, es_\alpha + {}^{u_2}\!/_{\sqrt{n}}; Y) - S(q_\alpha, es_\alpha; Y)\big) - u_2\,\partial_{x_2}S(q_\alpha, x_2; Y)\,(es_\alpha)\Big]$$

converges to zero in probability in $\big(\ell^\infty(K_2), \|\cdot\|_{K_2}\big)$. Using (3.11) and noting that

$$\partial_{x_2}S(q_\alpha, x_2; z)\,(es_\alpha) = G'(es_\alpha)\big(es_\alpha - q_\alpha + \alpha^{-1}\mathbb{1}\,(z \le q_\alpha)\,(q_\alpha - z)\big)$$

this implies

$$\begin{aligned}U_n(u_2) &= \sqrt{n}\,\mathbb{E}_n\Big[\sqrt{n}\big(S(q_\alpha, es_\alpha + {}^{u_2}\!/_{\sqrt{n}}; Y) - S(q_\alpha, es_\alpha; Y)\big)\Big] \\ &= u_2\sqrt{n}\mathbb{E}_n\big[G'(es_\alpha)\big(es_\alpha - q_\alpha + \tfrac{1}{\alpha}\mathbb{1}\,(Y \le q_\alpha)\,(q_\alpha - Y)\big)\big] + n\int_0^{\frac{u_2}{\sqrt{n}}} G'(es_\alpha + s)s\,\mathrm{d}s + o_{\mathrm{P}}(1) \\ &= u_2\,G'(es_\alpha)\frac{1}{\sqrt{n}}\sum_{i=1}^n\big(es_\alpha - q_\alpha + \tfrac{1}{\alpha}\mathbb{1}\,(Y_i \le q_\alpha)\,(q_\alpha - Y_i)\big) + \int_0^{u_2} G'(es_\alpha + \tfrac{t}{\sqrt{n}})t\,\mathrm{d}t + o_{\mathrm{P}}(1).\end{aligned} \tag{3.24}$$

Since the sequence $\left(es_\alpha - q_\alpha + \alpha^{-1}\mathbb{1}\left(Y_i \le q_\alpha\right)\left(q_\alpha - Y_i\right)\right)_{i\in\mathbb{N}}$ consists of centred, independent, and identically distributed random variables, using the central limit theorem, the first term in the last equality converges weakly in $\left(\ell^\infty(K_2), \|\cdot\|_{K_2}\right)$ to $G'(es_\alpha)\,u_2\,W_2$ for the stated W_2. The second term converges weakly in $\ell^\infty(K_2)$ equipped with the supremum distance to $\frac{1}{2}G'(es_\alpha)\,u_2^2$, and thus the limit process U of U_n has the asserted form.

To conclude the proof of (3.15), it remains to show the Lipschitz-property in (3.23). Suppose $K_2 \subset [-c_0, c_0]$ and choose $x_2, x_2' \in K_2$. Using (3.9) we compute

$$\begin{aligned}
&\left|S(q_\alpha, x_2'; z) - S(q_\alpha, x_2; z)\right| = \left|S(q_\alpha, x_2' + (x_2 - x_2'); z) - S(q_\alpha, x_2'; z)\right| \\
&= \left|\left(G(x_2) - G(x_2')\right)\left(x_2' - q_\alpha + \alpha^{-1}\mathbb{1}\left(z \le q_\alpha\right)\left(q_\alpha - z\right)\right) + \int_0^{x_2 - x_2'} G'(x_2' + s)s\,\mathrm{d}s\right| \\
&\le \left|G(x_2) - G(x_2')\right|\left(c_0 + |q_\alpha| + \alpha^{-1}\mathbb{1}\left(z \le q_\alpha\right)\left(q_\alpha - z\right)\right) + \left|\int_0^{x_2 - x_2'} G'(x_2' + s)s\,\mathrm{d}s\right|.
\end{aligned}$$

It follows from the mean value theorem that we can find a point $\xi \in K_2$ for which $\left|G(x_2) - G(x_2')\right| = \left|G'(\xi)(x_2 - x_2')\right|$. The right hand side of this is smaller than $C|x_2 - x_2'|$ as G' is continuous and hence bounded on K_2, such that $C = \sup_{x_2\in K_2} G'(x_2) < \infty$. This ends the discussion of the first addend above as we therefore obtain

$$\begin{aligned}
&\left|G(x_2) - G(x_2')\right|\left(c_0 + |q_\alpha| + \alpha^{-1}\mathbb{1}\left(z \le q_\alpha\right)\left(q_\alpha - z\right)\right) \\
&\qquad\qquad \le C\left(c_0 + |q_\alpha| + \alpha^{-1}\mathbb{1}\left(z \le q_\alpha\right)\left(q_\alpha - z\right)\right)|x_2 - x_2'|.
\end{aligned}$$

For the other addend we utilize the (second) mean value theorem to get

$$\begin{aligned}
\left|\int_0^{x_2 - x_2'} G'(x_2' + s)s\,\mathrm{d}s\right| &= \left|G'(x_2' + \xi)\xi(x_2 - x_2')\right| \qquad \text{for some } \xi \in [-2c_0, 2c_0] \\
&\le C_1 c_0 |x_2 - x_2'|
\end{aligned}$$

for $C_1 = \sup_{x_2\in[-3c_0, 3c_0]} G'(x_2)$, where $C_1 < \infty$ by continuity and hence boundedness of G' on $[-3c_0, 3c_0]$. All in all, we end up with

$$\left|S(q_\alpha, x_2'; z) - S(q_\alpha, x_2; z)\right| \le \left(C\left(c_0 + |q_\alpha| + \alpha^{-1}\mathbb{1}\left(z \le q_\alpha\right)\left(q_\alpha - z\right)\right) + C_1 c_0\right)|x_2 - x_2'|.$$

Denote the Lipschitz-constant on the right hand side with $L(z)$. Now, under Assumption [B] it holds that $\mathbb{E}\left[\mathbb{1}\left(Y \le 0\right)Y^2\right] < \infty$, so in this case $\mathbb{E}[L(Y)^2] < \infty$ is true. So (3.23) is indeed satisfied.

Finally, if Assumptions [A] and [B] are both valid, then the expansions (3.22) and (3.24) hold true and the joint process convergence follows, where the covariance of W_1 and W_2 is readily computed. □

The convergence of U_n proved beforehand is now used to conclude Step 4.

Proof of Lemma 3.13. The limit process U is quadratic and has the unique minimizer $u_2^0 = -W_2$. Further, from the form (3.14) of U_n and the argument leading to (3.2) it follows that the unique minimizer of U_n is given by

$$u_{2,n} = \sqrt{n}\,(\alpha^{-1}\mathbb{E}_n[\mathbb{1}\,(Y \le q_\alpha)\,(Y - q_\alpha)] - es_\alpha + q_\alpha)$$

which, using the central limit theorem, converges in distribution to $-W_2$ under Assumption [B].

To show (3.18) we apply Lemma 1.15 to the processes

$$M_n(u_2) = U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha + {}^{u_2}\!/\!\sqrt{n})\big),$$

and

$$M_n'(u_2) = U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha)\big),$$

so that U_n will play the role of N_n in Lemma 1.15, which converges weakly on compact sets equipped with the supremum norm to U by Lemma 3.16. Note that M_n is minimized by $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$; see (3.2). Now, in Lemma 3.11 we showed that $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$ is a tight sequence, and at the beginning of this proof we already showed weak convergence of the minimizers of U_n, namely $u_{2,n} \xrightarrow{\mathcal{L}} u_2^0$.

It thus remains to show that (1.9) holds true for the above choices of $M_n(u_2)$ and $M_n'(u_2)$, precisely

$$\begin{aligned}\sup_{u_2 \in K_2}\Big|&U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha + {}^{u_2}\!/\!\sqrt{n})\big)\\ &- \Big(U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha)\big)\Big)\Big| = o_{\mathrm{P}}(1).\end{aligned}$$

To this end, assume $K_2 \subset [-c_0, c_0]$. Then

$$\begin{aligned}\sup_{u_2 \in K_2}\Big|&U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha + {}^{u_2}\!/\!\sqrt{n})\big)\\ &- \Big(U_n(u_2) + V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\big(1 + \alpha^{-1}G(es_\alpha)\big)\Big)\Big|\\ &= \alpha^{-1}c_0\Big|\tfrac{1}{\sqrt{n}}V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\Big| \sup_{u_2 \in K_2}\left|\frac{G(es_\alpha + {}^{u_2}\!/\!\sqrt{n}) - G(es_\alpha)}{{}^{c_0}\!/\!\sqrt{n}}\right|\\ &\le \alpha^{-1}c_0\Big|\tfrac{1}{\sqrt{n}}V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha))\Big|\left|\frac{G(es_\alpha + {}^{c_0}\!/\!\sqrt{n}) - G(es_\alpha - {}^{c_0}\!/\!\sqrt{n})}{{}^{c_0}\!/\!\sqrt{n}}\right|\end{aligned}$$

is valid, since G is monotonically non-decreasing. The first two factors are constant, the last factor is $O(1)$ since the fraction converges to $2G'(es_\alpha)$, and it remains to show that $V_n\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha)\big) = o_{\mathrm{P}}(\sqrt{n})$, which would be implied by

$$V_n\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha)\big) = O_{\mathrm{P}}\big(\tfrac{\sqrt{n}}{a_n}\big). \tag{3.25}$$

To see (3.25), we first remark that $a_n(\widehat{q}_{n,\alpha} - q_\alpha)$ is a tight sequence. This follows from the results in Knight (2002), but is directly implied by (3.15), convexity of the V_n and V, and uniqueness of the minimizer of V with the aid of Lemma 2.2 in Davis et al. (1992). Thus, given $\varepsilon > 0$, there exists a compact set K_1 with $\mathrm{P}\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha) \in K_1\big) \geq 1 - \varepsilon$. Since for a fixed compact set K_1 the map $h \mapsto \inf_{K_1} h$, $h \in \ell^\infty(K_1)$, is continuous with respect to the supremum norm, (3.15) implies that $\inf_{K_1} \big(\frac{a_n}{\sqrt{n}}\big) V_n \xrightarrow{\mathcal{L}} \inf_{K_1} V$, in particular $\inf_{K_1} \big(\frac{a_n}{\sqrt{n}}\big) V_n$ is a tight sequence. To conclude, note that

$$\mathrm{P}\Big(\tfrac{a_n}{\sqrt{n}} V_n(a_n(\widehat{q}_{n,\alpha} - q_\alpha)) \geq C\Big) \leq \mathrm{P}\Big(\inf_{K_1} \tfrac{a_n}{\sqrt{n}} V_n \geq C\Big) + \mathrm{P}\big(a_n(\widehat{q}_{n,\alpha} - q_\alpha) \notin K_1\big),$$

which implies (3.25) and finishes the proof of the lemma. □

3.7 Remaining proofs

In this section we collect the remaining calculations and arguments. We start with the spared parts of Steps 1 and 2 in the proof of Theorem 3.5 and end with the proofs for the multi-dimensional and spectral version of Theorem 3.5.

3.7.1 Proofs for Step 1 and Step 2 of Theorem 3.5

First we show the assertions about the increments of the scoring function of (q_α, es_α).

Proof of Lemma 3.10. From (3.5), we have that

$$\begin{aligned}
&S(x_1 + y_1, x_2 + y_2; z) - S(x_1, x_2; z)\\
&= \big(1 + \alpha^{-1} G(x_2 + y_2)\big)\big(\mathbb{1}(z \leq x_1 + y_1) - \alpha\big)(x_1 + y_1 - z) + G(x_2 + y_2)(x_2 + y_2 - z)\\
&\quad - \mathcal{G}(x_2 + y_2) - \big(1 + \alpha^{-1} G(x_2)\big)\big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z) - G(x_2)(x_2 - z) + \mathcal{G}(x_2)\\
&= \big(1 + \alpha^{-1} G(x_2 + y_2)\big)\big(\mathbb{1}(z \leq x_1 + y_1) - \alpha\big)(x_1 + y_1 - z) - \big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z)\\
&\quad - \alpha^{-1} G(x_2)\big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z)\\
&\quad + G(x_2 + y_2)(x_2 + y_2 - z) - \mathcal{G}(x_2 + y_2) - G(x_2)(x_2 - z) + \mathcal{G}(x_2)\\
&= I) + II)
\end{aligned} \tag{3.26}$$

with

$$\begin{aligned}
I) &= \big(1 + \alpha^{-1} G(x_2 + y_2)\big)\big(\rho_\alpha(x_1 + y_1; z) - \rho_\alpha(x_1; z)\big),\\
II) &= \alpha^{-1} G(x_2 + y_2)\big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z) - \alpha^{-1} G(x_2)\big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z)\\
&\qquad + G(x_2 + y_2)(x_2 + y_2 - z) - \mathcal{G}(x_2 + y_2) - G(x_2)(x_2 - z) - \mathcal{G}(x_2),
\end{aligned}$$

where we subtracted and added the term $\big(\mathbb{1}(z \leq x_1) - \alpha\big)(x_1 - z)\alpha^{-1} G(x_2 + y_2)$ in (3.26).

Observe that $I) = 0$ when choosing $y_1 = 0$ and hence

$$S(x_1, x_2 + y_2; z) - S(x_1, x_2; z) = II) \tag{3.27}$$

is true, such that we have to further discuss $II)$. In $II)$ the terms $G(x_2 + y_2)z$ and $G(x_2)z$ cancel out, and rearranging gives

$$\begin{aligned}
II) &= G(x_2+y_2)\left(\alpha^{-1}\mathbb{1}\left(z \le x_1\right) - 1\right)x_1 - G(x_2)\left(\alpha^{-1}\mathbb{1}\left(z \le x_1\right) - 1\right)x_1 \\
&\quad - G(x_2+y_2)\alpha^{-1}\mathbb{1}\left(z \le x_1\right)z + G(x_2)\alpha^{-1}\mathbb{1}\left(z \le x_1\right)z \\
&\quad + G(x_2+y_2)x_2 - G(x_2)x_2 + G(x_2+y_2)y_2 - \mathcal{G}(x_2+y_2) + \mathcal{G}(x_2) \\
&= \left[\left(\alpha^{-1}\mathbb{1}\left(z \le x_1\right) - 1\right)x_1 - \alpha^{-1}\mathbb{1}\left(z \le x_1\right)z + x_2\right]\left(G(x_2+y_2) - G(x_2)\right) \\
&\quad + G(x_2+y_2)y_2 - \mathcal{G}(x_2+y_2) + \mathcal{G}(x_2) \\
&= \left(x_2 - x_1 + \alpha^{-1}\mathbb{1}\left(z \le x_1\right)\left(x_1 - z\right)\right)\left(G(x_2+y_2) - G(x_2)\right) \\
&\quad + G(x_2+y_2)y_2 - \mathcal{G}(x_2+y_2) + \mathcal{G}(x_2).
\end{aligned}$$

By a partial integration

$$G(x_2+y_2)y_2 - \mathcal{G}(x_2+y_2) + \mathcal{G}(x_2) = \int_0^{y_2} G'(x_2+s)s\,\mathrm{d}s$$

is valid, which together with (3.27) implies (3.9). A further partial integration gives (3.10).

To prove (3.11), note that by another partial integration,

$$\begin{aligned}
\rho_\alpha(x_1+y_1;z) - \rho_\alpha(x_1;z) &= \int_{x_1}^{x_1+y_1}\left(\mathbb{1}\left(z \le s\right) - \alpha\right)\mathrm{d}s = \int_0^{y_1}\left(\mathbb{1}\left(z \le x_1+s\right) - \alpha\right)\mathrm{d}s \\
&= -\alpha\, y_1 + \int_0^{y_1}\mathbb{1}\left(z \le x_1+s\right)\,\mathrm{d}s \\
&= y_1\left(\mathbb{1}\left(z \le x_1\right) - \alpha\right) + \int_0^{y_1}\left(\mathbb{1}\left(z \le x_1+s\right) - \mathbb{1}\left(z \le x_1\right)\right)\mathrm{d}s
\end{aligned}$$

where in the last equality we added and subtracted the term $y_1\mathbb{1}\left(z \le x_1\right)$. This is (3.11).

Finally, combining (3.26), (3.10), (3.27) and (3.11) gives (3.12). □

We already gave a sketch for the proof of Lemma 3.11. Now we present the detailed arguments.

Proof of Lemma 3.11. As indicated in the outline of the proof we use Theorem 1.14 with $\alpha = 2$, $\beta = 1$, d_0, d_1 the Euclidean distance in $\mathbb{R}$ and the criterion function $m_{\eta,\vartheta}(z) = S(\eta, \vartheta; z)$. The consistency of $\widehat{es}_{n,\alpha}$ was accomplished in Theorem 3.2.

Concerning (1.7) we need to prove that

$$\inf_{|\eta - q_\alpha| \le \varepsilon} \inf_{\delta_0 \ge |\vartheta - es_\alpha| \ge \delta} \mathbb{E}\left[S(\eta, \vartheta; Y) - S(\eta, es_\alpha; Y)\right] \ge C\,\delta^2 \tag{3.28}$$

is valid for all $0 < \varepsilon \le \varepsilon_0$, $0 < \delta \le \delta_0$ and some $C, \delta_0, \varepsilon_0 > 0$.

To this end, using (3.9) and $\rho_\alpha(x; y)$ in Lemma 3.10, we get that

$$\begin{aligned}
&\inf_{|\eta - q_\alpha| \le \varepsilon} \inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \mathbb{E}\left[S(\eta, es_\alpha + \varepsilon_2; Y) - S(\eta, es_\alpha; Y)\right] \\
&= \inf_{|\eta - q_\alpha| \le \varepsilon} \inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \Bigg[(G(es_\alpha + \varepsilon_2) - G(es_\alpha)) \mathbb{E}\left[es_\alpha - \eta + \tfrac{1}{\alpha} \mathbb{1}\left(Y \le \eta\right)(\eta - Y)\right] \\
&\qquad\qquad + \int_0^{\varepsilon_2} G'(es_\alpha + s) s \,\mathrm{d}s \Bigg] \\
&= \inf_{|\eta - q_\alpha| \le \varepsilon} \inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \Bigg[(G(es_\alpha + \varepsilon_2) - G(es_\alpha)) \mathbb{E}\left[es_\alpha - Y + \tfrac{1}{\alpha}\Big(\left(\mathbb{1}\left(Y \le \eta\right) - \alpha\right)(\eta - Y)\Big)\right] \\
&\qquad\qquad + \int_0^{\varepsilon_2} G'(es_\alpha + s) s \,\mathrm{d}s \Bigg], \\
&\ge \inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \left(G(es_\alpha + \varepsilon_2) - G(es_\alpha)\right) \inf_{|\eta - q_\alpha| \le \varepsilon} \left(es_\alpha - \mathbb{E}\left[Y\right] + \tfrac{1}{\alpha} \mathbb{E}\left[\rho_\alpha(\eta; Y)\right]\right) \\
&\quad + \inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \int_0^{\varepsilon_2} G'(es_\alpha + s) s \,\mathrm{d}s.
\end{aligned}$$

The function $\eta \mapsto \rho_\alpha(\eta; F)$ attains its (unique) minimum in q_α, as it is a strictly consistent scoring function for the α-quantile; see Example 1.7. But $es_\alpha - \mathbb{E}[Y] + \alpha^{-1}\rho_\alpha(q_\alpha; F) = 0$ and thus the expression in the last inequality above is greater than (or equal to)

$$\inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \int_0^{\varepsilon_2} G'(es_\alpha + s) s \,\mathrm{d}s.$$

The remaining integral is monotonically increasing for $\varepsilon_2 > 0$ and decreasing for $\varepsilon_2 < 0$, therefore the infimum is attained in $\pm\delta$. A partial integration then gives

$$\begin{aligned}
\inf_{\delta_0 \ge |\varepsilon_2| \ge \delta} \int_0^{\varepsilon_2} G'(es_\alpha + s) s \,\mathrm{d}s &= \delta^2 \left(\frac{1}{2} G'(es_\alpha \pm \delta) - \delta^{-2} \frac{1}{2} \int_0^{\pm\delta} G''(es_\alpha + s) s^2 \,\mathrm{d}s \right) \\
&\ge \frac{\delta^2}{2} \left(G'(es_\alpha \pm \delta) - \delta^{-2} C_1 \int_0^{\pm\delta} s^2 \,\mathrm{d}s \right) \\
&\ge \frac{\delta^2}{2} \Big(G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3} \Big)
\end{aligned}$$

for

$$C_1 = \sup_{x_2 \in [es_\alpha - \delta_0, es_\alpha + \delta_0]} |G''(x_2)| < \infty.$$

Since $\lim_{\delta \to 0} G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3} = G'(es_\alpha) > 0$ holds by assumption on G' we can find $C_2 > 0$ and $\delta_0 > 0$ with $G'(es_\alpha \pm \delta) - \frac{\delta\, C_1}{3} \geq C_2$ for every $\delta \leq \delta_0$. This proves (3.28) with $C = \frac{C_2}{2}$.

Next, (1.8) translates to

$$\mathbb{E}\Bigg[\sup_{\substack{|\eta - q_\alpha| \leq \varepsilon \\ |\vartheta - es_\alpha| \leq \delta}} \Big|\mathbb{G}_n\big[S(\eta, \vartheta; Y) - S(\eta, es_\alpha; Y)\big]\Big|\Bigg] \leq C\,\delta \tag{3.29}$$

for any $0 < \delta \leq \delta_0$, and some $\delta_0, C > 0$. To see this inequality we use (3.9) again and the fact that the increment $G(es_\alpha + \varepsilon_2) - G(es_\alpha)$ equals $\int_0^{\varepsilon_2} G'(es_\alpha + s)\,\mathrm{d}s$ to obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{|\eta - q_\alpha| \leq \varepsilon} \sup_{|\varepsilon_2| \leq \delta} \big|\mathbb{G}_n\big[S(\eta, es_\alpha + \varepsilon_2; Y) - S(\eta, es_\alpha; Y)\big]\big|\right] \\
&= \mathbb{E}\left[\sup_{|\eta - q_\alpha| \leq \varepsilon} \sup_{|\varepsilon_2| \leq \delta} \Big| \int_0^{\varepsilon_2} G'(es_\alpha + s)\,\mathrm{d}s\, \mathbb{G}_n\big[\alpha^{-1} \mathbb{1}\,(Y \leq \eta)\,(\eta - Y)\big]\Big|\right]
\end{aligned}$$

(only the stochastic term remains). Since $G' > 0$, the former expression does not exceed

$$\int_{-\delta}^{\delta} G'(es_\alpha + s)\,\mathrm{d}s\, \alpha^{-1} \mathbb{E}\left[\sup_{|\eta - q_\alpha| \leq \varepsilon} \big|\mathbb{G}_n[\mathbb{1}\,(Y \leq \eta)\,(\eta - Y)]\big|\right].$$

Because the first integral fulfils

$$\int_{-\delta}^{\delta} G'(es_\alpha + s)\,\mathrm{d}s \leq 2\,\delta G'(\xi)$$

for some $\xi \in [es_\alpha - \delta_0, es_\alpha + \delta_0]$ by the mean value theorem and $G' > 0$, it is sufficient to show

$$\mathbb{E}\left[\sup_{|\eta - q_\alpha| \leq \varepsilon} \big|\mathbb{G}_n[\mathbb{1}\,(Y \leq \eta)\,(\eta - Y)]\big|\right] \leq C \tag{3.30}$$

for some constant C not depending on δ.

To this end we use a maximal inequality involving the bracketing integral as introduced in Definition 2.6. Observe for any $\eta \in [q_\alpha - \varepsilon, q_\alpha + \varepsilon]$ the inequality

$$\big|\mathbb{1}\,(z \leq \eta)\,(\eta - z)\big| \leq \mathbb{1}\,(z \leq q_\alpha + \varepsilon)\,(q_\alpha + \varepsilon - z).$$

Thus, $H(z) = \mathbb{1}\,(z \le q_\alpha + \varepsilon)\,(q_\alpha + \varepsilon - z)$ is an envelope function for the (measurable) class of functions $\mathcal{H} = \{z \mapsto \mathbb{1}\,(z \le \eta)\,(\eta - z)\,|\,\eta \in [q_\alpha - \varepsilon, q_\alpha + \varepsilon]\}$. Using Corollary 19.35, van der Vaart (1998), we obtain

$$\mathbb{E}\left[\sup_{|\eta - q_\alpha| \le \varepsilon} \big|\mathbb{G}_n[\mathbb{1}\,(Y \le \eta)\,(\eta - Y)]\big|\right] \le C_1 \int_0^{C_2} \sqrt{\log N_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})}\,\mathrm{d}\delta$$

for some constant $C_1 < \infty$ and $C_2 = \|H\|_{Y,2}$, where $N_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})$ denotes the bracketing number with respect to the norm $\|f\|_{Y,2} = \big(\mathbb{E}[f(Y)^2]\big)^{1/2}$ in Definition 2.6; note that $C_2 < \infty$ is true under Assumption [B]. Next, observe that the class $\mathcal{H}$ fulfils a Lipschitz-condition, namely for any $\eta_1, \eta_2 \in [q_\alpha - \varepsilon, q_\alpha + \varepsilon]$ it holds that

$$\big|\mathbb{1}\,(z \le \eta_1)\,(\eta_1 - z) - \mathbb{1}\,(z \le \eta_2)\,(\eta_2 - z)\big| \le |\eta_1 - \eta_2|.$$

As seen in Example 19.7, van der Vaart (1998), there is a constant C_3 only depending on ε, such that the bracketing number satisfies

$$N_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2}) \le \frac{2\,C_3\,\varepsilon}{\delta}$$

for any $0 < \delta < 2\,\varepsilon$. Hence, by partitioning the bracketing integral we are left with

$$\begin{aligned}
\int_0^{C_2} \sqrt{\log N_{[\,]}(\delta, \mathcal{H}, \|\cdot\|_{Y,2})}\,\mathrm{d}\delta &\le \int_0^{2\,\varepsilon} \sqrt{\log\big(\tfrac{2\,C_3\,\varepsilon}{\delta}\big)}\,\mathrm{d}\delta + C_4 \\
&\le \sqrt{2\,C_3\,\varepsilon} \int_0^{2\,\varepsilon} \delta^{-1/2}\,\mathrm{d}\delta + C_4 \\
&= 4\,\varepsilon\,\sqrt{C_3} + C_4
\end{aligned}$$

for a constant C_4 not depending on δ. Putting things together we have shown

$$\mathbb{E}\left[\sup_{|\eta - q_\alpha| \le \varepsilon} \big|\mathbb{G}_n[\mathbb{1}\,(Y \le \eta)\,(\eta - Y)]\big|\right] \le C_1\big(4\,\sqrt{C_3}\,\varepsilon + C_4\big),$$

what is (3.30) by choosing $C = C_1\big(4\,\sqrt{C_3}\,\varepsilon + C_4\big)$. □

3.7.2 Proofs of Theorems 3.8 and 3.9

Here, we show the multi-dimensional and spectral versions of Theorem 3.5. For the proof of Theorem 3.8, we again need the processes V_n and U_n used while proving Theorem 3.5.

Proof of Theorem 3.8. We define the processes V_n^s and U_n^s as in (3.13) and (3.14) for each α_s, $s = 1, \ldots, k$. Then the expansions (3.22) and (3.24) are valid for each s, and the covariance matrix in the joint normal distribution of the $2k$-dimensional vector $(W_{1,1}, W_{1,2}, \dots, W_{k,1}, W_{k,2})$ in the limit processes U^s and V^s, which are given by

$$V^s(u_1) = u_1 W_{s,1} + \int_0^{u_1} \psi_{\alpha_s}(t)\,\mathrm{d}t \text{ and } U^s(u_2) = G'(es_{\alpha_s})\,\big(\,u_2\, W_{s,2} + \tfrac{u_2^2}{2}\big),$$

see Lemma 3.12, is determined by

$$\mathrm{Cov}(W_{s,1}, W_{t,2}) = \mathbb{E}\big[\big(\mathbb{1}\,(Y \le q_{\alpha_s}) - \alpha_s\big)\big(\mathbb{1}\,(Y \le q_{\alpha_t}) - \alpha_t\big)\big] = \alpha_s \wedge \alpha_t - \alpha_s\alpha_t,$$

$$\begin{aligned}
&\mathrm{Cov}(W_{s,2}, W_{t,2})\\
&= \mathbb{E}\Big[\big(es_{\alpha_s} - q_{\alpha_s} + \alpha_s^{-1}\mathbb{1}\,(Y \le q_{\alpha_s})\,(q_{\alpha_s} - Y)\big)\big(es_{\alpha_t} - q_{\alpha_t} + \alpha_t^{-1}\mathbb{1}\,(Y \le q_{\alpha_t})\,(q_{\alpha_t} - Y)\big)\Big]\\
&= (\alpha_t\,\alpha_s)^{-1}\mathbb{E}\Big[\mathbb{1}\,(Y \le q_{\alpha_s} \wedge q_{\alpha_t})\,(q_{\alpha_s} - Y)(q_{\alpha_t} - Y)\Big] \;+ (es_{\alpha_s} - q_{\alpha_s})(es_{\alpha_t} - q_{\alpha_t})\\
&= \frac{\alpha_s \wedge \alpha_t}{\alpha_s\alpha_t}\big(q_{\alpha_s}q_{\alpha_t} - (q_{\alpha_s} + q_{\alpha_t})es_{\alpha_s\wedge\alpha_t}\big) + (\alpha_t\,\alpha_s)^{-1}\mathbb{E}\big[\mathbb{1}\,(Y \le q_{\alpha_s} \wedge q_{\alpha_t})\,Y^2\big]\\
&\quad + (es_{\alpha_s} - q_{\alpha_s})(es_{\alpha_t} - q_{\alpha_t})
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Cov}(W_{s,1}, W_{t,2}) &= \mathbb{E}\Big[\big(\mathbb{1}\,(Y \le q_{\alpha_s}) - \alpha_s\big)\big(es_{\alpha_t} - q_{\alpha_t} + \alpha_t^{-1}\mathbb{1}\,(Y \le q_{\alpha_t})\,(q_{\alpha_t} - Y)\big)\Big]\\
&= \alpha_t^{-1}\mathbb{E}\Big[\big(\mathbb{1}\,(Y \le q_{\alpha_s}) - \alpha_s\big)(q_{\alpha_t} - Y)\mathbb{1}\,(Y \le q_{\alpha_t})\Big]\\
&= \frac{\alpha_t \wedge \alpha_s}{\alpha_t}(q_{\alpha_t} - es_{\alpha_s\wedge\alpha_t}) - \alpha_s(q_{\alpha_t} - es_{\alpha_t}),
\end{aligned}$$

where $s, t \in \{1, \dots, k\}$. Further, Lemma 3.13 also holds true for each s. As in Step 5 of the proof of Theorem 3.5, we may then consider the sequence of minimizers of the processes

$$Z_{n,mult}(v_1, u_1, \dots, v_k, u_k) = \sum_{s=1}^{k} \Big[\big(1 + \alpha_s^{-1}G(es_{\alpha_s})\big)\frac{a_{s,n}}{\sqrt{n}}V_n^s(v_s) + U_n^s(u_s)\Big],$$

which converge weakly in $\big(\ell^\infty(K), \|\cdot\|_K\big)$, $K \subset \mathbb{R}^{2k}$ compact, to the process

$$Z_{mult}(v_1, u_1, \dots, v_k, u_k) = \sum_{s=1}^{k} \Big[\big(1 + \alpha_s^{-1}G(es_{\alpha_s})\big)V^s(v_s) + U^s(u_s)\Big],$$

and apply the argmax-continuity theorem to obtain the result. □

Last, we show the assertions about the spectral risk measure κ_m, where the important part is the representation of $\widehat{\kappa}_{m,n}$ in (3.4).

Proof of Theorem 3.9. Concerning (3.4) we work as in the case of the Expected Shortfall (see Lemma 3.1), and set

$$g_s(x_{k+1}) = 1 + \frac{p_s}{\alpha_s} G(x_{k+1}), \quad h(x_{k+1}; z) = G(x_{k+1})(x_{k+1} - z) - \mathcal{G}(x_{k+1}).$$

Then we have that

$$(\widehat{q}_{n,\alpha_1}, \ldots, \widehat{q}_{n,\alpha_k}, \widehat{\kappa}_{m,n}) \in \operatorname*{arg\,min}_{x_1,\ldots,x_{k+1}\in\mathbb{R}} \sum_{s=1}^{k} \sum_{i=1}^{n} \Big[g_s(x_{k+1}) \rho_{\alpha_s}(x_s; Y_i) + p_s\, h(x_{k+1}; Y_i) \Big].$$

The minimal value equals

$$\min_{x_{k+1}\in\mathbb{R}} \sum_{s=1}^{k} \Big[g_s(x_{k+1}) \Big(\min_{x_s\in\mathbb{R}} \sum_{i=1}^{n} \rho_{\alpha_s}(x_s; Y_i) \Big) + p_s \sum_{i=1}^{n} h(x_{k+1}; Y_i) \Big],$$

so the minimizer in x_s does not depend on x_l, $l \neq s$, and is actually given by $\widehat{q}_{n,\alpha_s}$. It remains to find the minimizer of the function

$$x_{k+1} \mapsto \sum_{s=1}^{k} \Big[\sum_{i=1}^{n} g_s(x_{k+1}) \rho_{\alpha_s}(\widehat{q}_{n,\alpha_s}; Y_i) + p_s\, h(x_{k+1}; Y_i) \Big].$$

Differentiation of the maps g_s and h gives

$$\partial_{x_{k+1}} g_s(x_{k+1}) = \frac{p_s}{\alpha_s} G'(x_{k+1}), \quad \partial_{x_{k+1}} h(x_{k+1}; z) = G'(x_{k+1})(x_{k+1} - z),$$

so that minimizing the above function is equivalent to solving

$$0 = G'(x_{k+1}) \sum_{s=1}^{k} p_s \Big[n\, x_{k+1} - n\, \widehat{q}_{n,\alpha_s} + \alpha_s^{-1} \sum_{i=1}^{n} \mathbb{1}\left(Y_i \leq \widehat{q}_{n,\alpha_s}\right) \left(\widehat{q}_{n,\alpha_s} - Y_i\right) \Big]$$

for x_{k+1}, which results in

$$\widehat{\kappa}_{m,n} = \sum_{s=1}^{k} p_s \left[\widehat{q}_{n,\alpha_s} - \frac{1}{n\,\alpha_s} \sum_{i=1}^{n} \mathbb{1}\left(Y_i \leq \widehat{q}_{n,\alpha_s}\right) \left(\widehat{q}_{n,\alpha_s} - Y_i\right) \right].$$

Utilizing formula (3.2) for the Expected Shortfall then implies (3.4).

Using Theorem 3.8 in combination with the continuous mapping theorem then shows

$$\sqrt{n}\left(\widehat{\kappa}_{m,n} - \kappa_m\right) \xrightarrow{\mathcal{L}} \sum_{s=1}^{k} p_s\, W_{s,2}$$

as stated. □

Chapter 4

Considerations for general Bayes risks

In this chapter we deduced the joint weak convergence of a vector of elicitable parameters together with the respective Bayes risk by generalizing the ideas of Chapter 3.

4.1 Introduction

The pair (Value at Risk, Expected Shortfall) is a special case of a $k+1$-dimensional parameter, where the first entries are given by an elicitable parameter with score S and the last entry is the Bayes risk with respect to the fixed score S. Other important examples are the pairs (Mean, Variance) and (Expectile, Variantile) or the vector $(q_{\alpha_1}, \ldots, q_{\alpha_k}, \kappa_m)$ considered in the former chapter. Another example is the bivariate risk measure (Mean, g(Mean)), where g is a strictly convex function with $\mathbb{E}\left[g(Y)\right] < \infty$.

This chapter generalises the arguments of Chapter 3. Here, we indicate how to obtain the joint asymptotic distribution of (T, γ), where the k-dimensional parameter T is elicitable with respect to some class $\mathcal{F}$ of distribution functions and $\gamma(F) = -\min_{x \in \mathbb{R}^k} S_0(x; F)$ for some strictly $\mathcal{F}$-consistent score S_0 for T.

The structure of the chapter is similar to the former one. In Section 4.2 we define the parameter and estimator, for which we establish results concerning consistency and weak convergence in Section 4.3. The proofs are relegated to Section 4.5.

4.2 Preliminaries

Let $\mathcal{F}$ be a family of distribution functions and $T : \mathcal{F} \to \mathbb{R}^k$, $k \in \mathbb{N}$, be a parameter which is elicitable with respect to $\mathcal{F}$. We choose a strictly $\mathcal{F}$-consistent scoring function S_0

for T and define the functional $\gamma_0 : \mathcal{F} \to \mathbb{R}$, $\gamma_0(F) = -\min_{x_1 \in \mathbb{R}^k} S_0(x_1; F) = S_0(T; F)$. We are interested in the $k+1$-dimensional parameter $T_0 = (T, \gamma_0)$. By Theorem 1.9 the property T_0 is elicitable with strictly consistent scoring function S given by

$$S(x_1, x_2; z) = S_0(x_1; z) + G(x_2)\big(x_2 + S_0(x_1; z)\big) - \big(\mathcal{G}(x_2) - \mathcal{G}(z)\big),$$

where $\mathcal{G}$ is three-times continuously differentiable with $\mathcal{G}' = G$ and G is strictly increasing with $\lim_{x \to -\infty} G(x) = 0$.

Assuming existence, we define the M-estimators $(T_n, \widehat{\gamma}_n)$ as the empirical counterpart coming from an independent identically distributed sequence of random variables $Y, Y_1, \ldots, Y_n$ with $Y, Y_i \sim F$, where $F \in \mathcal{F}$, namely $(T_n, \widehat{\gamma}_n) \in \arg\min S(x_1, x_2; F_n)$.

We want to examine the behaviour of $(T_n, \widehat{\gamma}_n)$ as an estimator for (T, γ_0) with respect to consistency and asymptotic distribution. As in the case for the Expected Shortfall we concentrate on the assumptions needed to ensue consistency of $\widehat{\gamma}_n$ and existence of an asymptotic distribution thereof, when already knowing these properties for T_n.

4.3 Joint asymptotic theory for T_n and $\widehat{\gamma}_n$

In this section we show consistency of $(T_n, \widehat{\gamma}_n)$ and derive the asymptotic distribution of a centred and properly rescaled version thereof. We will state exemplary conditions for which the results do hold as well as indicate further possibilities.

4.3.1 Consistency of $\widehat{\gamma}_n$

Let us start with the consistency of $\widehat{\gamma}_n$. The following calculation is similar to the proof of Theorem 1.9.

First, by the subgradient inequality for $\mathcal{G}$, any $x_2 \neq -S_0(x_1; F_n)$ fulfils

$$\begin{aligned} S\big(x_1, -S_0(x_1; F_n); F_n\big) &= S_0(x_1; F_n) - \big(\mathcal{G}(-S_0(x_1; F_n)) - \mathcal{G}(F_n)\big) \\ &< S_0(x_1; F_n) + G(x_2)\big(x_2 + S_0(x_1; F_n)\big) - \big(\mathcal{G}(x_2) - \mathcal{G}(F_n)\big) \\ &= S(x_1, x_2; F_n), \end{aligned}$$

so that the minimum of $x_2 \mapsto S(x_1, x_2; F_n)$ is uniquely determined by $x_2^* = -S_0(x_1; F_n)$. Letting $\widetilde{S}(x; F_n) = S(x_1, x_2^*; F_n)$, we have

$$\arg\min_{x_1 \in \mathbb{R}^k} \widetilde{S}(x_1; F_n) = \arg\min_{x_1 \in \mathbb{R}^k} \Big(S_0(x_1; F_n)\Big) \cap \arg\min_{x_1 \in \mathbb{R}^k} \Big(-\mathcal{G}(-S_0(x_1; F_n))\Big).$$

As $\mathcal{G}$ is strictly increasing, this reduces to

$$\begin{aligned} \arg\min_{x_1 \in \mathbb{R}^k} \widetilde{S}(x_1; F_n) &= \arg\min_{x_1 \in \mathbb{R}^k} \Big(S_0(x_1; F_n)\Big) \cap \arg\max_{x_1 \in \mathbb{R}^k} \Big(-S_0(x_1; F_n)\Big) \\ &= \arg\min_{x_1 \in \mathbb{R}^k} \Big(S_0(x_1; F_n)\Big). \end{aligned}$$

Thus, it follows that

$$\underset{x_1\in\mathbb{R}^k,\, x_2\in\mathbb{R}}{\arg\min}\ S(x_1,x_2;F_n) = \Big\{\big(t,\ -S_0(t;F_n)\big) \ \big|\ t\in\underset{x_1\in\mathbb{R}^k}{\arg\min}\,\big(S_0(x_1;F_n)\big)\Big\}. \tag{4.1}$$

Next, observe that by positivity of G for any x_2 it holds that

$$\begin{aligned} T_n &\in \underset{x_1\in\mathbb{R}^k}{\arg\min}\, S(x_1,x_2;F_n) \\ &= \underset{x_1\in\mathbb{R}^k}{\arg\min}\, S_0(x_1;F_n) + G(x_2)\big(x_2 + S_0(x_1;F_n)\big) - \big(\mathcal{G}(x_2) - \mathcal{G}(F_n)\big) \\ &= \underset{x_1\in\mathbb{R}^k}{\arg\min}\,\big(1+G(x_2)\big)\, S_0(x_1;F_n) = \underset{x_1\in\mathbb{R}^k}{\arg\min}\, S_0(x_1;F_n). \end{aligned}$$

Hence, $t = T_n$ is a possible choice in (4.1), for which $\widehat{\gamma}_n = -S_0(T_n;F_n)$ is valid. So the consistency of $(T_n,\widehat{\gamma}_n)$ has to come from the assumptions on S_0 for which the following is one possibility. The proof is deferred to Section 4.5.

4.1 Theorem.

Assume that T_n is consistent for T and $x\mapsto S_0(x_1;z)$ is continuous. Further, suppose that for some $\varepsilon>0$ the class of functions $\mathcal{H}_\varepsilon = \{z\mapsto S_0(x_1;z)\,|\,x_1\in\mathrm{cl}(B_\varepsilon(T))\}$ is Glivenko-Cantelli. Then $\widehat{\gamma}_n\to\gamma_0$ in probability.

For consistency of T_n, which is also defined as an M-estimator, there are several assumptions present in the literature. For example, we can assume the whole class $\{z\mapsto S_0(x_1;z)\,|\,x_1\in\mathbb{R}^k\}$ to be Glivenko-Cantelli (van der Vaart, 1998, Theorem 5.7). There are weaker assumptions, for example, when knowing that T_n lies in a compact set $K\subset\mathbb{R}^k$ eventually (van der Vaart, 1998, Theorem 5.14). Another possibility is to assume asymptotic stochastic equicontinuity of $x_1\mapsto S_0(x_1;F_n) - S_0(x_1;F)$. This and more alternatives can be found in Newey and McFadden (1994); see their discussion after Theorem 2.1.

Next, we consider the asymptotic distribution of $\big(a_n(T_n - T), b_n(\widehat{\gamma}_n - \gamma_0)\big)$ where $a_n, b_n\to\infty$ are deterministic sequences.

4.3.2 Rate of convergence for $\widehat{\gamma}_n$

We argue that $\sqrt{n}(\widehat{\gamma}_n - \gamma_0) = O_{\mathrm{P}}(1)$ if we impose appropriate assumptions on S_0.

For this we want to utilize Theorem 1.14 where we need to show that for fixed C, every $n\in\mathbb{N}$ and all sufficiently small $\varepsilon,\delta>0$ the inequalities

$$\inf_{|\eta-T|\le\varepsilon}\ \inf_{\delta_0\ge|\varepsilon_2|\ge\delta} \mathbb{E}\big[S(\eta,\gamma_0+\varepsilon_2;Y) - S(\eta,\gamma_0;Y)\big] \ge C\delta^2 \tag{4.2}$$

and

$$\mathbb{E}^\circ\Big[\sup_{|\eta-T|\le\varepsilon}\sup_{\delta_0\ge|\varepsilon_2|\ge\delta}\big|\mathbb{G}_n\big[S(\eta,\gamma_0+\varepsilon_2;Y)-S(\eta,\gamma_0;Y)\big]\big|\Big]\le C\delta \tag{4.3}$$

are valid. Therefore we calculate the increments for the scoring function S. The proof is analogue to the one of Lemma 3.10; we omit the details.

4.2 Lemma.

For any $x_1, y_1 \in \mathbb{R}^k$ and $x_2, y_2 \in \mathbb{R}$ it holds that

$$\begin{aligned}
&S(x_1,x_2+y_2;z)-S(x_1,x_2;z)\\
&=\big(G(x_2+y_2)-G(x_2)\big)\big(x_2+S_0(x_1;z)\big)+\int_0^{y_2}G'(x_2+s)s\,\mathrm{d}s\\
&=\big(G(x_2+y_2)-G(x_2)\big)\big(x_2+S_0(x_1;z)\big)+\tfrac12 G'(x_2+y_2)y_2^2-\tfrac12\int_0^{y_2}G''(x_2+s)s^2\,\mathrm{d}s.
\end{aligned} \tag{4.4}$$

Increments in the first coordinate satisfy

$$S(x_1+y_1,x_2;z)-S(x_1,x_2;z)=\big(1+G(x_2)\big)\left(S_0(x_1+y_1;z)-S_0(x_1;z)\right).$$

Especially, it holds that

$$\begin{aligned}
&S(x_1+y_1,x_2+y_2;z)-S(x_1,x_2;z)\\
&\quad=\big(1+G(x_2+y_2)\big)\big(S_0(x_1+y_1;z)-S_0(x_1;z)\big)\\
&\qquad+\big(G(x_2+y_2)-G(x_2)\big)\big(x_2+S_0(x_1;z)\big)\\
&\qquad+\frac12 G'(x_2+y_2)y_2^2-\frac12\int_0^{y_2}G''(x_2+s)s^2\,\mathrm{d}s.
\end{aligned} \tag{4.5}$$

With the aid of (4.4) and consistency of S_0 for T we can estimate (4.2) as

$$\begin{aligned}
&\inf_{|\eta-T|\le\varepsilon}\inf_{\delta_0\ge|\varepsilon_2|\ge\delta}\mathbb{E}\big[S(\eta,\gamma_0+\varepsilon_2;Y)-S(\eta,\gamma_0;Y)\big]\\
&=\inf_{|\eta-T|\le\varepsilon}\inf_{\delta_0\ge|\varepsilon_2|\ge\delta}\Big[\left(G(\gamma_0+\varepsilon_2)-G(\gamma_0)\right)\left(\gamma_0+\mathbb{E}\left[S_0(\eta;Y)\right]\right)+\int_0^{\varepsilon_2}G'(\gamma_0+s)\,s\,\mathrm{d}s\Big]\\
&\ge\inf_{\delta_0\ge|\varepsilon_2|\ge\delta}\Big[\left(G(\gamma_0+\varepsilon_2)-G(\gamma_0)\right)\left(\gamma_0+\inf_{|\eta-T|\le\varepsilon}\mathbb{E}\left[S_0(\eta;Y)\right]\right)\\
&\qquad\qquad+\inf_{\delta_0\ge|\varepsilon_2|\ge\delta}\int_0^{\varepsilon_2}G'(\gamma_0+s)\,s\,\mathrm{d}s\Big]\\
&=\inf_{\delta_0\ge|\varepsilon_2|\ge\delta}\int_0^{\varepsilon_2}G'(\gamma_0+s)\,s\,\mathrm{d}s.
\end{aligned}$$

This is greater than (or equal to) $C\,\delta^2$ when choosing δ_0 small enough as seen in the proof of (3.28).

So it remains to show (4.3), for which the following assertion is one possibility.

4.3 Theorem.

Inequality (4.3) *is valid, provided for some* $C > 0$ *it holds that*

$$\mathbb{E}\left[\sup_{|\eta - T| \leq \varepsilon} \left|\mathbb{G}_n[S_0(\eta; Y)]\right|\right] \leq C.$$

In Lemma 3.11 a maximal inequality using the bracketing integral is applied to obtain the constant C needed in the former theorem. This can be done here as well under appropriate conditions on S_0. For example, it is sufficient to assume that $\mathcal{H}_\varepsilon$ introduced in Theorem 4.1 has an envelope function H with $\mathbb{E}\left[H(Y)^2\right] < \infty$ and a bracketing integral fulfilling

$$J_{[\,]}(\mathbb{E}[H(Y)^2], \mathcal{H}_\varepsilon, \|\cdot\|_{Y,2}) < \infty.$$

This would also cover $\mathcal{H}_\varepsilon$ being Glivenko-Cantelli (van der Vaart, 1998, Theorem 19.5) as needed for consistency in Theorem 4.1. For the boundedness of the bracketing integral we could assume $\mathcal{H}_\varepsilon$ to be a Lipschitz class of functions, namely

$$\left|S_0(\eta; z) - S_0(\eta'; z)\right| \leq L(z)\,|\eta - \eta'|$$

for a measurable function $L(z)$ fulfilling $\mathbb{E}\left[L(Y)^2\right] < \infty$; see van der Vaart (1998), Example 19.7.

So far we have shown the following.

4.4 Theorem.

Let the assumption of Theorem 4.3 be true. Then $\sqrt{n}\,(\widehat{\gamma}_n - \gamma_0) = O_{\mathrm{P}}(1)$ *is valid.*

4.3.3 Asymptotic distribution

Let us assume $b_n = \sqrt{n}$ for the rest of this chapter, meaning $\sqrt{n}\,(\widehat{\gamma}_n - \gamma_0) = O_{\mathrm{P}}(1)$ holds; for example, we use the assumptions of Theorem 4.3. Next, we determine the joint asymptotic distribution of $\big(a_n(T_n - T), \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big)$. As in the proof of Theorem 3.5, we therefore utilize the weak convergence of the (rescaled) processes

$$V_n(u) = \sum_{i=1}^{n} S_0(T + {}^{u}/_{a_n}; Y_i) - S_0(T; Y_i),$$
$$U_n(u) = \sum_{i=1}^{n} S\big(T, \gamma_0 + {}^{u}/_{\sqrt{n}}; Y_i\big) - S\big(T, \gamma_0; Y_i\big)$$

to some limit processes V and U with unique minimizers in order to show weak convergence of $\big(a_n\,(T_n - T),\, \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big)$. Note that the processes V_n and U_n fulfil

$$\begin{aligned}
&\big(1 + G(\gamma_0 + {}^{u_2}/_{\sqrt{n}})\big)\, V_n(u_1) + U_n(u_2) \\
&= \big(1 + G(\gamma_0 + {}^{u_2}/_{\sqrt{n}})\big) \sum_{i=1}^{n} \Big(S_0(T + {}^{u_1}/_{a_n}; Y_i) - S_0(T; Y_i)\Big) \\
&\quad + \sqrt{n}\,\big(G(\gamma_0 + {}^{u_2}/_{\sqrt{n}}) - G(\gamma_0)\big)\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\gamma_0 + S_0(T; Y_i)\big) + \frac{u_2^2}{2}\, G'(\gamma_0 + {}^{u_2}/_{\sqrt{n}}) \\
&\quad - \frac{1}{2\sqrt{n}} \int_0^{u_2} G''(\gamma_0 + {}^{s}/_{\sqrt{n}})\, s^2 \,\mathrm{d}s \\
&= n\,\big(S_0\big(T + {}^{u_1}/_{a_n}, \gamma_0 + {}^{u_2}/_{\sqrt{n}}; F_n\big) - S_0\big(T, \gamma_0; F_n\big)\big),
\end{aligned}$$

see (4.5). The value $\big(a_n(T_n - T), \sqrt{n}(\widehat{\gamma}_n - \gamma_0)\big)$ is a minimizer of the above process; moreover $a_n\big(T_n - T\big) \in \arg\min_u V_n(u)$ is valid.

As before, V_n and U_n can have different rates of convergence. In the spirit of the proof of Theorem 3.5, approximating $\widehat{\gamma}_n$ by the minimizer of U_n, which we call $u_{2,n}$, does help in this situation. For this approximation we use Lemma 1.15. In order for that approach to work, we first need weak convergence of U_n to some limit process U and existence of a unique minimizer of U.

4.5 Lemma.

If $\mathbb{E}\left[S_0(T; Y)^2\right] < \infty$, *then*

$$\begin{aligned}
U_n(u_2) &= u_2\, G'(\gamma_0)\, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\gamma_0 + S_0(T; Y_i)\big) + \int_0^{u_2} G'(\gamma_0 + {}^{s}/_{\sqrt{n}})\, s \,\mathrm{d}s + o_{\mathrm{P}}(1) \\
&= u_2\, G'(\gamma_0)\, \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big(\gamma_0 + S_0(T; Y_i)\big) + G'\big(\gamma_0 + {}^{u_2}/_{\sqrt{n}}\big)\, \frac{u_2^2}{2} + o_{\mathrm{P}}(1)
\end{aligned}$$

is valid. Here the $o_{\mathrm{P}}(1)$*-term is a sequence of processes in* $\big(\ell^\infty(K_2), \|\cdot\|_{K_2}\big)$*, where* $K_2 \subset \mathbb{R}$ *is compact.*

This immediately yields the weak convergence of U_n.

4.6 Corollary.

Assume that $\mathbb{E}\left[S_0(T;Y)^2\right] < \infty$ *and let* $K_2 \subset \mathbb{R}$ *be compact. Then* U_n *converges in distribution in the space* $\left(\ell^\infty(K_2), \|\cdot\|_{K_2}\right)$ *to the process* U *given by*

$$U(u_2) = G'(\gamma_0)\big(u_2\, W_2 + \tfrac{u_2^2}{2}\big),$$

where $W_2 \sim \mathcal{N}(0, \mathbb{E}\left[S_0(T;Y)^2\right] - \gamma_0^2)$*. Especially,* U_n *converges in distribution to a limit process with unique minimizer which is given by* $u_2^0 = -W_2$*.*

Note that in addition, U_n itself has a unique minimizer, determined as

$$u_{2,n} = \sqrt{n}\,\big(-S_0(T;F_n) - \gamma_0\big),$$

what can be seen by the same calculations that lead to Theorem 4.1. This representation implies $u_{2,n} \xrightarrow{\mathcal{L}} u_2^0$ with the aid of the central limit theorem, provided we assume $\mathbb{E}\left[S_0(T,Y)^2\right] < \infty$.

Next, we show the approximation $u_{2,n} = \sqrt{n}(\widehat{\gamma}_n - \gamma_0) + o_{\mathrm{P}}(1)$, for which we observe the representation

$$u_{2,n} = \sqrt{n}\,\big(-S_0(T;F_n) - \gamma_0\big) = \sqrt{n}(\widehat{\gamma}_n - \gamma_0) + \sqrt{n}\big(S_0(T_n;F_n) - S_0(T;F_n)\big).$$

The second summand converges to 0 in probability if we assume, for example, a Lipschitz-condition on S_0 and consistency of T_n. An alternative is the following.

4.7 Theorem.

Assume that $r_n V_n \rightsquigarrow V$ *in* $\left(\ell^\infty(K_1), \|\cdot\|_{K_1}\right)$ *for any compact set* $K_1 \subset \mathbb{R}^k$ *and a sequence* $r_n \to \infty$*. Further, suppose that* $a_n(T_n - T) = O_{\mathrm{P}}(1)$ *and* $\mathbb{E}\left[S_0(T;Y)^2\right] < \infty$*. Then* $u_{2,n} = \sqrt{n}(\widehat{\gamma}_n - \gamma_0) + o_{\mathrm{P}}(1)$ *is true.*

The former theorem implies

$$\big(a_n\,(T_n - T),\, \sqrt{n}(\widehat{\gamma}_n - \gamma_0)\big) = \big(a_n\,(T_n - T),\, u_{2,n}\big) + o_{\mathrm{P}}(1).$$

Now, the variable $\big(a_n\,(T_n - T),\, u_{2,n}\big)$ is a minimizer of the process $V_n(u_1) + U_n(u_2)$ by construction. Further, as the variables are separated, it is also a minimizer of the process

$$Z_n(u_1, u_2) = \big(1 + G(\gamma_0)\big)\, r_n\, V_n(u_1) + U_n(u_2). \tag{4.6}$$

We therefore can formulate the following assertion about the joint asymptotic distribution of $\big(a_n\,(T_n - T),\, \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big)$.

4.8 Theorem.

Assume that $r_n V_n \rightsquigarrow V$ *in* $\big(\ell^\infty(K_1), \|\cdot\|_{K_1}\big)$ *for any compact set* $K_1 \subset \mathbb{R}^k$ *where* V *almost surely has a unique minimizer* v_0 *and continuous sample paths. In addition, suppose that* $a_n (T_n - T) = O_{\mathrm{P}}(1)$ *and* $\mathbb{E}\left[S_0(T;Y)^2\right] < \infty$. *Then the sequence* $\big(a_n (T_n - T), \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big)$ *converges in distribution to* (v_0, u_2^0).

4.4 Conclusion and discussion

In this chapter we showed a possible extension of the scheme used in Chapter 3. We imposed more or less minimal requirements, such that the idea of the proof of Theorem 3.5 still works, and indicated certain assumptions on S_0 and T_n which guarantee the validity of Theorem 4.8. Under these assumptions we saw that the rate of convergence of $\widehat{\gamma}_n$ is not influenced by the rate of convergence of T_n.

A next possible extension is a multivariate version of Theorem 4.8, where the course for a proof can be seen in the proof of Theorem 3.8.

4.5 Proofs

This section comprises the proofs of the statements in the present chapter. They are largely parallel to the proofs in Section 3.6.

We start with proving the consistency of $\widehat{\gamma}_n$.

Proof of Theorem 4.1. Note that

$$\widehat{\gamma}_n = -S_0(T_n;F) + \big(S_0(T_n;F) - S_0(T_n;F_n)\big)$$

is valid. The bracketed term converges to zero in probability as follows. Observe that for any $\varepsilon, \delta > 0$ the inequality

$$\mathrm{P}\Big(\big|S_0(T_n;F) - S_0(T_n;F_n)\big| > \delta\Big) \le \mathrm{P}\Big(\sup_{x\in[T-\varepsilon,T+\varepsilon]} \big|S_0(x;F) - S_0(x;F_n)\big| > \delta\Big) + \mathrm{P}\big(|T_n - T| > \varepsilon\big)$$

holds. Both probabilities on the right hand side converge to zero for $n \to \infty$: The first by the assumption on $\mathcal{H}_\varepsilon$ being Glivenko-Cantelli, the second by assumed consistency of T_n for T. Thus, it holds that $\widehat{\gamma}_n = -S_0(T_n;F) + o_{\mathrm{P}}(1)$, and we need to discuss $-S_0(T_n;F)$.

Therefore observe that, as $x \mapsto S_0(x;z)$ is continuous, the map $x \mapsto S_0(x;F)$ is continuous as well, for example with the aid of Fubini's Theorem. Hence, as T_n converges to T in probability, it follows that $-S_0(T_n;F)$ converges in probability to $-S_0(T;F) = \gamma_0$ using the continuous mapping theorem. $\square$

In order to obtain the rate of convergence of $\widehat{\gamma}_n$, it remained to show (1.8) of Theorem 1.14.

Proof of Theorem 4.3. We apply the same reductions as in the proof of Lemma 3.11 starting below (3.29). After this it remains to bound $\mathbb{E}\big[\sup_{|\nu-T|\le\varepsilon}|\mathbb{G}[S_0(\nu;Y)]|\big]$, which is smaller than some constant C by assertion. $\square$

Lemma 4.5 gives an asymptotic expansion of U_n. The proof is similar to Lemma 3.12 and uses Lemma 19.31, van der Vaart (1998).

Proof of Lemma 4.5. We show a Lipschitz-property to be valid. Let K_2 be a compact set. Using the representation in (4.4) we can estimate

$$\begin{aligned}
&|S(x_1,x_2;z)) - S(x_1,x_2';z))|\\
&= |S(x_1,x_2'+(x_2-x_2');z) - S(x_1,x_2';z)|\\
&\le \big|(G(x_2)-G(x_2'))\,(x_2'+S_0(x_1;z))\big| + \Big|\int_0^{x_2-x_2'} G'(x_2'+s)\,s\,\mathrm{d}s\Big|\\
&\le \sup_{\xi\in K_2} G'(\xi)|x_2-x_2'|\,(\sup_{\xi\in K_2}|\xi| + |S_0(x_1;z)|) + \sup_{\xi\in 3\,K_2} G'(x_2+\xi)\,|\xi|\,|x_2-x_2'|\\
&= |x_2-x_2'|\Big(\sup_{\xi\in K_2} G'(\xi)\,(\sup_{\xi\in K_2}|\xi| + |S_0(x_1;z)|) + \sup_{\xi\in 3\,K_2} G'(x_2+\xi)\,|\xi|\Big).
\end{aligned}$$

Let $L(z)$ be the Lipschitz constant occurring in the last equality. As $\mathbb{E}\left[S_0(T;Y)^2\right]<\infty$, it follows that $\mathbb{E}\left[L(Y)^2\right]<\infty$. Next, note that $x_2\mapsto S(x_1,x_2;z)$ is differentiable with derivative

$$\partial_{x_2} S(x_1,x_2;z) = G'(x_2)\,(x_2+S_0(x_1;z)).$$

Hence, Lemma 19.31, van der Vaart (1998), yields

$$\mathbb{G}\Big[\sqrt{n}(S(T,\gamma_0+{}^{u_2}\!/\!_{\sqrt{n}};Y) - S(T,\gamma_0;Y)) - u_2\,G'(\gamma_0)\,(\gamma_0+S_0(T;Y))\Big] = o_{\mathrm{P}}(1), \quad (4.7)$$

where the $o_{\mathrm{P}}(1)$-term is a sequence of processes in $(\ell^\infty(K_2),\|\cdot\|_{K_2})$. As the equality

$$U_n(u_2) = \sqrt{n}\,\mathbb{E}_n\left[\sqrt{n}\left(S(T,\gamma_0+{}^{u_2}\!/\!_{\sqrt{n}};Y) - S(T,\gamma_0;Y)\right)\right]$$

is valid, we can therefore conclude by reorganizing (4.7). More detailed, we get

$$\begin{aligned}
U_n(u_2) &= \sqrt{n}\,\mathbb{E}_n\left[\sqrt{n}\left(S(T,\gamma_0+{}^{u_2}\!/\!_{\sqrt{n}};Y) - S(T,\gamma_0;Y)\right)\right]\\
&= u_2\sqrt{n}\,\mathbb{E}_n\left[G'(\gamma_0)\,(\gamma_0+S_0(T;Y))\right] + n\Big(\int_0^{\frac{u_2}{\sqrt{n}}} G'(T+s)\,s\,\mathrm{d}s\Big) + o_{\mathrm{P}}(1)\\
&= u_2\sqrt{n}\,\mathbb{E}_n\left[G'(\gamma_0)\,(\gamma_0+S_0(T;Y))\right] + \int_0^{u_2} G'(T+{}^{s}\!/\!_{\sqrt{n}})\,s\,\mathrm{d}s + o_{\mathrm{P}}(1)\\
&= u_2\,G'(\gamma_0)\,\sqrt{n}\,\mathbb{E}_n\left[\gamma_0+S_0(T;Y)\right] + G'(\gamma_0+{}^{u_2}\!/\!_{\sqrt{n}})\,\frac{u_2^2}{2} + o_{\mathrm{P}}(1).
\end{aligned}$$

$\square$

Using this representation we can show weak convergence of U_n.

Proof of Corollary 4.6. By Lemma 4.5 we know that

$$U_n(u_2) = u_2\, G'(\gamma_0)\, \sqrt{n}\, \mathbb{E}_n\left[\gamma_0 + S_0(T;Y)\right] + G'\big(\gamma_0 + {}^{u_2}\!/\!\sqrt{n}\big)\, \frac{u_2^2}{2} + o_{\mathrm{P}}(1)$$

and with the aid of Slutsky's lemma we only have to deal with the weak convergence of the first two summands. The first term converges weakly to $u_2 \mapsto u_2 G'(\gamma_0) W_2$ for W_2 as asserted; the second converges to the map $u_2 \mapsto G'(\gamma_0)\frac{u_2^2}{2}$, where both convergences hold in $\big(\ell^\infty(K_2), \|\cdot\|_{K_2}\big)$. This shows $U_n \rightsquigarrow U$.

Last, as $G' > 0$, minimizing $U(u_2)$ is equivalent to minimizing $u_2 W_2 + {}^{u_2^2}\!/\!{}_2$, which yields the asserted minimizer $u_2^0 = -W_2$. □

In Theorem 4.7 we approximate $\widehat{\gamma}_n$ with the minimizers of U_n, for which we use Lemma 1.15.

Proof of Theorem 4.7. In Corollary 4.6 we have seen that U_n converges weakly on compact sets to U with respect to the supremum distance. Further, we showed that U_n has a unique minimizer $u_{2,n}$, which converges in distribution to the unique minimizer u_2^0 of U (see below Corollary 4.6). In addition, we assumed $a_n(T_n - T) = O_{\mathrm{P}}(1)$, such that it remains to show

$$\begin{aligned}\sup_{u \in K_2} \Big| U_n(u) &+ V_n\big(a_n(T_n - T)\big)\left(1 + G(\gamma_0 + \tfrac{u}{\sqrt{n}})\right) \\ &- U_n(u) - V_n\big(a_n(T_n - T)\big)\left(1 + G(\gamma_0)\right)\Big| = o_{\mathrm{P}}(1)\end{aligned}$$

for any compact K_2 in order to apply Lemma 1.15. Therefore assume $K_2 \subset [-c_0, c_0]$, then

$$\begin{aligned}\sup_{u \in K_2} \Big| U_n(u) &+ V_n\big(a_n(T_n - T)\big)\left(1 + G(\gamma_0 + \tfrac{u}{\sqrt{n}})\right) \\ &- U_n(u) - V_n\big(a_n(T_n - T)\big)\left(1 + G(\gamma_0)\right)\Big| \\ &= \frac{c_0}{\sqrt{n}} \left|V_n\big(a_n(T_n - T)\big)\right| \sup_{u \in K_2} \left| \frac{G(\gamma_0 + {}^{u}\!/\!\sqrt{n}) - G(\gamma_0)}{{}^{c_0}\!/\!\sqrt{n}} \right| \\ &\leq \frac{c_0}{r_n\sqrt{n}} \left|r_n\, V_n\big(a_n(T_n - T)\big)\right| \sup_{u \in K_2} \left| \frac{G(\gamma_0 + {}^{c_0}\!/\!\sqrt{n}) - G(\gamma_0 - {}^{c_0}\!/\!\sqrt{n})}{{}^{c_0}\!/\!\sqrt{n}} \right|\end{aligned}$$

is true due to monotonicity of G. Note that the first fraction converges to zero whereas the last fraction converges to $2\, G'(\gamma_0)$ and thus is bounded. Hence, it remains to prove

$$r_n\, V_n\big(a_n(T_n - T)\big) = O_{\mathrm{P}}(1). \tag{4.8}$$

By assumption, $a_n(T_n - T)$ is a tight sequence such that for given $\varepsilon > 0$ there exists a compact set K_1 with $\mathrm{P}\big(a_n(T_n - T) \in K_1\big) \geq 1 - \varepsilon$. Now, the map $h \mapsto \inf_{K_1} h$ is continuous for $h \in \ell^\infty(K_1)$ with respect to the supremum metric, and thus the weak convergence $\inf_{K_1} r_n V_n \xrightarrow{\mathcal{L}} \inf_{K_1} V$ follows, using the weak convergence $r_n V_n \rightsquigarrow V$. In particular, $\inf_{K_1} r_n V_n$ is a tight sequence.

Last, we observe the inequality

$$\mathrm{P}\Big(r_n V_n\big(a_n(T_n - T)\big) \geq C\Big) \leq \mathrm{P}\Big(\inf_{K_1} r_n V_n \geq C\Big) + \mathrm{P}\big(a_n(T_n - T) \notin K_1\big),$$

where both probabilities on the right hand side can be made small. This shows (4.8) and concludes the proof. □

For the next proof, which determines the asymptotic distribution of the random variable $\big(a_n(T_n - T),\, \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big)$, we utilize the argmax-continuity theorem, Corollary 5.58, van der Vaart (1998), as done in Step 5 of the proof of Theorem 3.5.

Proof of Theorem 4.8. The sequence of processes Z_n in (4.6) converges weakly by assumption, Corollary 4.6 and the continuous mapping theorem to the process

$$Z(u_1, u_2) = \big(1 + G(\gamma_0)\big)\, V(u_1) + U(u_2),$$

where the convergence takes place in $\big(\ell^\infty(K), \|\cdot\|_K\big)$ for any compact set $K \subset \mathbb{R}^{k+1}$. This process almost surely has a unique minimizer given by (v_0, u_2^0); further, it almost surely has continuous sample paths by assumption on V and the shape of U in Corollary 4.6. Next, the sequence $u_{2,n}$ is tight by Corollary 4.6 and the discussion thereafter, and the sequence $a_n(T_n - T)$ is tight by assumption. Thus, the sequence $\big(a_n(T_n - T), u_{2,n}\big)$, which minimizes Z_n, is tight. Hence, the argmax-continuity theorem yields the weak convergence $\big(a_n(T_n - T), u_{2,n}\big) \xrightarrow{\mathcal{L}} (v_0, u_2^0)$. Finally, by Theorem 4.7, we know that $\big(a_n(T_n - T),\, \sqrt{n}\,(\widehat{\gamma}_n - \gamma_0)\big) = \big(a_n(T_n - T),\, u_{2,n}\big) + o_{\mathrm{P}}(1)$ is valid, which implies the assertion. □

Chapter 5

Process convergence of expectiles and quantiles

In this chapter we showed weak convergence of the empirical quantile and expectile processes to Gaussian limit processes in the space of bounded functions endowed with the hypi-semimetric under general assumptions on the underlying distribution function. We further showed how to obtain convergence in the M_2-topology directly, without using the equivalence to the hypi-topology. We also considered the validity of the bootstrap and examined the results with a numerical illustration.

5.1 Introduction

Quantiles are fundamental parameters of a probability distribution which have various applications in statistics and econometrics (Koenker, 2005) as well as in finance (McNeil et al., 2015). For distributions with finite first moments, expectiles are defined as minimizers of a weighted quadratic loss and have found renewed interest as a coherent elicitable class of risk measures (Bellini, Klar, et al., 2014; Ziegel, 2016). See Chapter 1 for further details.

The asymptotic properties of sample quantiles and expectiles have been addressed in detail under suitable conditions. For quantiles, differentiability of the distribution function at the quantile with positive derivative implies asymptotic normality of the empirical quantile; see Chapter 3. In addition, under a continuity assumption on the density we obtain weak convergence of the quantile process to a Gaussian limit process in the space of bounded functions endowed with the supremum distance from the functional delta-method (van der Vaart, 1998). However, without the existence of a positive derivative of the underlying distribution function at the quantile, the weak (pointwise) limit will be non-normal (see Theorem 3.5), and thus process convergence to a Gaussian limit with respect to the supremum distance cannot hold true.

Similarly, for a distribution with finite second moment, the empirical expectile is asymptotically normally distributed if the distribution function is continuous at the expectile but non-normally distributed otherwise (Holzmann and Klar, 2016). For continuous distribution functions, process convergence of the empirical expectile process in the space of continuous functions equipped with the supremum topology also holds true, but for discontinuous distribution functions this can no longer be valid.

In this chapter we discuss convergence of quantile and expectile processes from independent and identically distributed observations under more general conditions. Indeed, we show that the expectile process converges to a Gaussian limit in the semimetric space of bounded functions endowed with the hypi-semimetric as introduced in Section 2.3.1 under the assumption of a finite second moment only. Since the Gaussian limit process is discontinuous in general while the empirical expectile process is continuous, this convergence cannot hold with respect to the supremum distance. As we will see the hypi-semimetric is appropriate in this situation; due to the equivalence of the topologies, the M_2-topology works as well.

Further, we show weak convergence of the quantile process under the hypi-semimetric if the distribution function is admits finitely many point masses and has left- and right-sided derivatives – which can be infinite – in every point, which are bounded from below.

These results still imply weak convergence of important statistics such as Kolmogorov-Smirnov and Cramér-von Mises type statistics. Moreover, we show consistency of the n-out-of-n bootstrap in both situations.

The track for this chapter is the following. In Section 5.2 we state the limit results mentioned above. Section 5.4 contains a short simulation study, which in particular illustrates the discontinuity of the limit process. We conclude and indicate further extensions in Section 5.5. Section 5.6 contains an outline of the proofs of the main results as well as details for the most relevant steps. Section 5.8 carries the results over to Skorohod M_2-convergence while proofs of technical assertions are deferred to Section 5.9.

5.2 Weak convergence of quantile and expectile processes

Here we formulate the main results of this chapter regarding the weak convergence of the empirical expectile process and the empirical quantile process, both with respect to the hypi-semimetric. We argue that the assumptions are quite weak and boundedness from below of the one-sided derivatives seems to be necessary for the result for quantiles. In addition, we link the results to the assertions of Section 2.3.2, precisely we elaborate that the results are in fact applications of Theorem 2.15 and Lemma 2.16.

5.2.1 Convergence of the expectile process

For a random variable Y with distribution function $F \in \mathcal{L}_1$, the τ-expectile $\mu_\tau = \mu_\tau(F)$, $\tau \in (0,1)$, in Definition 1.10 is the minimizer of an expected score. Alternatively, μ_τ can be defined as the unique solution of $\mathbb{E}\left[I_\tau(x;Y)\right] = 0$, $x \in \mathbb{R}$, where

$$I_\tau(x;z) = \tau(z-x)\mathbb{1}\left(z \geq x\right) - \left(1-\tau\right)\left(x-z\right)\mathbb{1}\left(z < x\right).$$

Given a sequence of independent and identically distributed copies $Y_1, Y_2, \ldots$ of Y and a natural number $n \in \mathbb{N}$, we let

$$\hat{\mu}_{\tau,n} = \mu_\tau(F_n)$$

be the empirical τ-expectile. Our main result for the expectile process is the following.

5.1 Theorem.

Suppose that $\mathbb{E}\left[Y^2\right] < \infty$. Given $0 < \tau_l < \tau_u < 1$ such that F is continuous in $\mu_{\tau_l}, \mu_{\tau_u}$, the standardized expectile process $\tau \mapsto \sqrt{n}\big(\hat{\mu}_{\tau,n} - \mu_\tau\big)$, $\tau \in [\tau_l, \tau_u]$, converges weakly in $\big(\ell^\infty([\tau_l,\tau_u]), \mathrm{d}_{hypi}\big)$ to the limit process $\big(\dot{\psi}_0^{\mathrm{Inv}}(Z)(\tau)\big)_{\tau\in[\tau_l,\tau_u]}$. Here,

$$\dot{\psi}_0^{\mathrm{Inv}}(\varphi)(\tau) = \varphi(\tau)\left(\tau + (1-2\,\tau)F(\mu_\tau)\right)^{-1}, \qquad \varphi \in \ell^\infty[\tau_l,\tau_u],$$

and $(Z_\tau)_{\tau\in[\tau_l,\tau_u]}$ is a centred tight Gaussian process with continuous sample paths and covariance function $\mathrm{Cov}(Z_\tau, Z_{\tau'}) = \mathbb{E}\left[I_\tau(\mu_\tau;Y)\, I_{\tau'}(\mu_{\tau'};Y)\right]$ for $\tau, \tau' \in [\tau_l, \tau_u]$.

Note that if F in the former theorem is continuous in a neighbourhood of $[\tau_l, \tau_u]$, the limit process $\dot{\psi}_0^{\mathrm{Inv}}(Z)$ almost surely has continuous sample paths, such that the hypi-convergence of the empirical expectile process implies convergence in the supremum distance thereof; see the discussion after Proposition 2.12. Thus, Theorem 5.1 is a generalisation of Theorem 8, Holzmann and Klar (2016).

From Propositions 2.3 and 2.4 in Bücher et al. (2014), hypi-convergence of the expectile process implies ordinary weak convergence of important statistics such as Kolmogorov-Smirnov or Cramér-von Mises type statistics.

5.2 Corollary.

If $\mathbb{E}\left[Y^2\right] < \infty$ and F is continuous in $\mu_{\tau_l}, \mu_{\tau_u}$, we have as $n \to \infty$ that

$$\sqrt{n}\|\widehat{\mu}_{\cdot,n} - \mu_\cdot\|_{[\tau_l,\tau_u]} \xrightarrow{\mathcal{L}} \|\dot{\psi}_0^{\mathrm{Inv}}(Z)\|_{[\tau_l,\tau_u]}.$$

Further, for $p \geq 1$ and a bounded non-negative weight function w on $[\tau_l, \tau_u]$,

$$n^{p/2} \int_{\tau_l}^{\tau_u} |\widehat{\mu}_{\tau,n} - \mu_\tau|^p \, w(\tau) \, \mathrm{d}\tau \xrightarrow{\mathcal{L}} \int_{\tau_l}^{\tau_u} |\dot{\psi}_0^{\mathrm{Inv}}(Z)(\tau)|^p \, w(\tau) \, \mathrm{d}\tau.$$

5.3 Remark (Point evaluation).
Evaluation at a given point x is only a continuous operation under the hypi-semimetric if the limit function is continuous at x; see Proposition 2.2 in Bücher et al. (2014). In particular, this does not apply to the expectile process if the distribution function F is discontinuous at μ_τ. Indeed, Theorem 7 in Holzmann and Klar (2016) shows that the weak limit of the empirical expectile is not normal in this case. $\diamond$

Next, we turn to the validity of the bootstrap. Given $n \in \mathbb{N}$ let $Y_1^*, \ldots, Y_n^*$ denote a sample drawn from $Y_1, \ldots, Y_n$ with replacement, that is, having distribution function F_n. Let F_n^* denote the empirical distribution function of $Y_1^*, \ldots, Y_n^*$, and let $\mu_{\tau,n}^* = \mu_\tau(F_n^*)$ denote the bootstrap expectile at level $\tau \in (0, 1)$.

5.4 Theorem.

Let the assumptions of Theorem 5.1 be true. Then, almost surely, conditionally on $Y_1, Y_2, \ldots$ the standardized bootstrap expectile process $\tau \mapsto \sqrt{n}(\mu_{\tau,n}^* - \hat{\mu}_{\tau,n})$, $\tau \in [\tau_l, \tau_u]$, converges weakly in $(\ell^\infty([\tau_l, \tau_u]), \mathrm{d}_{hypi})$ to $\big(\dot{\psi}_0^{\mathrm{Inv}}(Z)(\tau)\big)_{\tau \in [\tau_l, \tau_u]}$, where the map $\dot{\psi}_0^{\mathrm{Inv}}$ and the process $(Z_\tau)_{\tau \in [\tau_l, \tau_u]}$ are as in Theorem 5.1.

The simple n-out-of-n bootstrap does not apply for the empirical expectile at level τ if F is discontinuous at μ_τ, see Knight (1998) for a closely related result for the quantile. Thus, Theorem 5.4 is somewhat surprising, but its conclusion is reasonable in view of Remark 5.3.

5.2.2 Convergence of the quantile process

Let $q_\alpha = q_\alpha(F)$ and denote with $\widehat{q}_{\alpha,n} = q_\alpha(F_n)$ the empirical α-quantile of the sample $Y_1, \ldots, Y_n$. Concerning the distribution function F we assume the following.

Assumption [C].
For given $0 < \alpha_l < \alpha_u < 1$ and $\varepsilon > 0$, the distribution function F is strictly increasing on $[q_{\alpha_l} - \varepsilon, q_{\alpha_u} + \varepsilon]$. Assume that F is continuous except for finitely many points $\{y_1, \ldots, y_r\}$. In addition, assume that F admits right- and left-sided derivatives – which may be

infinite – at any point of $(q_{\alpha_l} - \varepsilon, q_{\alpha_u} + \varepsilon)$, that is

$$\partial^+(F)(q) = \lim_{h \to 0, h>0} \frac{F(q+h) - F(q)}{h} \quad \text{and} \quad \partial^-(F)(q) = \lim_{h \to 0, h>0} \frac{F(q) - F(q-h)}{h}$$

exist in $\mathbb{R} \cup \{\infty\}$ for any $q \in (q_{\alpha_l} - \varepsilon, q_{\alpha_u} + \varepsilon)$. Further, suppose that both functions $q \mapsto \partial^+(F)(q)$ and $q \mapsto \partial^-(F)(q)$ are bounded from below by some $c > 0$, are càdlàg or làdcàg in every point except for $\{y_1, \ldots, y_r\}$, and continuous in α_l and α_u. $\star$

First, note that strict monotonicity of F is necessary for consistency of the empirical quantile; see Koenker (2005), Section 4.1.1. An absolutely continuous example for which the derivative is unbounded can be obtained by glueing together in $1/4$ the function $x \mapsto -\sqrt{-x + 1/4} + 1/2$, $x \in [0, 1/4]$, and a normal distribution function with mean $1/4$ and variance 1. However, a continuous singular distribution function as, for example, the Cantor distribution does not satisfy Assumption [C].

Under Assumption [C] we are able to determine the left- and right-sided derivatives of F^{Inv} which are important for our main theorem.

5.5 Lemma.

Let Assumption [C] hold for the distribution function F. Then the map $\alpha \mapsto F^{\mathrm{Inv}}(\alpha)$ admits right- and left-sided derivatives $\partial^+(F^{\mathrm{Inv}})$ and $\partial^-(F^{\mathrm{Inv}})$ in every point of the interval $(\alpha_l - \delta, \alpha_u + \delta)$ for some $\delta > 0$. Choose any $\alpha \in (\alpha_l - \delta, \alpha_u + \delta)$.

i) If F is continuous in q_α, the derivatives are given by

$$\partial^-(F^{\mathrm{Inv}})(\alpha) = \left(\partial^-(F)(q_\alpha)\right)^{-1} \quad \text{and} \quad \partial^+(F^{\mathrm{Inv}})(\alpha) = \left(\partial^+(F)(q_\alpha)\right)^{-1}.$$

Now, let F jump in q_α.

ii) If $\alpha \in \left(F(q_\alpha-), F(q_\alpha)\right)$ the derivatives are determined by $\partial^\pm(F^{\mathrm{Inv}})(\alpha) = 0$.

iii) If $\alpha = F(q_\alpha)$, it holds that

$$\partial^-(F^{\mathrm{Inv}})(\alpha) = 0 \quad \text{and} \quad \partial^+(F^{\mathrm{Inv}})(\alpha) = \left(\partial^+(F)(q_\alpha)\right)^{-1}.$$

iv) If $\alpha = F(q_\alpha-)$, they are given by

$$\partial^-(F^{\mathrm{Inv}})(\alpha) = \left(\partial^-(F)(q_\alpha-)\right)^{-1} \text{ and } \quad \partial^+(F^{\mathrm{Inv}})(\alpha) = 0.$$

Especially, the maps $\alpha \mapsto \partial^-(F^{\mathrm{Inv}})(\alpha)$ and $\alpha \mapsto \partial^+(F^{\mathrm{Inv}})(\alpha)$ are càdlàg or làdcàg in every point.

With this we can formulate the following result for the empirical quantile process.

5.6 Theorem.

Given $0 < \alpha_l < \alpha_u < 1$ suppose that the distribution function F satisfies Assumption [C]. Then the standardised quantile process $\alpha \mapsto \sqrt{n}(\widehat{q}_{\alpha,n} - q_\alpha)$, $\alpha \in [\alpha_l, \alpha_u]$, converges weakly in $(\ell^\infty([\alpha_l, \alpha_u]), \mathrm{d}_{hypi})$ to the process $\big(\partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha\big)_{\alpha\in[\alpha_l,\alpha_u]}$, where $(V_\alpha)_{\alpha\in[\alpha_l,\alpha_u]}$ is a Brownian bridge on $[\alpha_l, \alpha_u]$.
Furthermore, as $n \to \infty$ we have that

$$\sqrt{n}\,\big\|\widehat{q}_{\cdot,n} - q_\cdot\big\|_{[\alpha_l,\alpha_u]} \xrightarrow{\mathcal{L}} \big\|\partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha\big\|_{[\alpha_l,\alpha_u]}, \tag{5.1}$$

as well as

$$n^{p/2} \int_{\alpha_l}^{\alpha_u} \big|\widehat{q}_{\alpha,n} - q_\alpha\big|^p\, w(\alpha)\,\mathrm{d}\,\alpha \xrightarrow{\mathcal{L}} \int_{\alpha_l}^{\alpha_u} \big|\partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha\big|^p\, w(\alpha)\,\mathrm{d}\alpha$$

for any $p \geq 1$ and a bounded, non-negative weight function w on $[\alpha_l, \alpha_u]$.

Here we observe that if F fulfils the standard assumption for the asymptotic behaviour of the empirical quantile process, namely that F is continuously differentiable on a neighbourhood of $[q_{\alpha_l}, q_{\alpha_u}]$ with strictly positive derivative f, F also fulfils Assumption [C] and the results of Theorem 5.6 hold true. In that case, the function $\partial^-(F^{\mathrm{Inv}})$ reduces to $\partial^-(F^{\mathrm{Inv}})(\alpha) = f(q_\alpha)$, and hence the limit process $\big(\partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha\big)_\alpha$ is almost surely continuous. Thus, the convergence of the empirical quantile process does hold with respect to the supremum distance, showing that Theorem 5.6 generalizes Example 3.9.24, van der Vaart and Wellner (1996).

Now, for $n \in \mathbb{N}$, let $Y_1^*, \ldots, Y_n^*$ again denote a sample drawn from $Y_1, \ldots, Y_n$ with replacement, that is, having distribution function F_n. As above use F_n^* to denote the empirical distribution function of $Y_1^*, \ldots, Y_n^*$ and let $q_{\alpha,n}^* = q_\alpha(F_n^*)$ denote the bootstrap quantile at level $\alpha \in (0,1)$.

5.7 Theorem.

Let the assumptions of Theorem 5.6 hold. Then, the standardized bootstrap quantile process $\alpha \mapsto \sqrt{n}(q_{\alpha,n}^* - \widehat{q}_{\alpha,n})$, $\alpha \in [\alpha_l, \alpha_u]$, converges weakly in $\big(\ell^\infty([\alpha_l, \alpha_u]), \mathrm{d}_{hypi}\big)$ to $\big(\partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha\big)_{\alpha\in[\alpha_l,\alpha_u]}$ conditionally on $Y_1, Y_2, \ldots$ in probability.

5.8 Remark (Boundedness in Assumption [C]).
The conclusion (5.1) of Theorem 5.6 in particular implies that $\sqrt{n}\|\widehat{q}_{n,\cdot} - q_\cdot\| = O_{\mathrm{P}}(1)$.

However, if for some quantile q_{α_0}, $\alpha_0 \in (\alpha_l, \alpha_u)$, the one-sided derivative of F satisfies $\partial^+(F)(q_{\alpha_0}) = 0$ or $\partial^-(F)(q_{\alpha_0}) = 0$, then the discussion in Example 3.4 shows that for a sequence a_n with $\frac{a_n}{\sqrt{n}} \to 0$, $a_n\,(\widehat{q}_{n,\alpha_0} - q_{\alpha_0})$ converges in distribution to a degenerate limit, a contradiction (see also Knight (2002)). ◇

5.2.3 Connection to Theorem 2.15 and Lemma 2.16

In this section we connect the statements of Theorems 5.1 and 5.6 to the general framework of Section 2.3.2. In doing this, we also suggest further weakening of the assumptions in the respective theorems.

5.9 Example (Quantile process).
As we will see, Theorem 5.6 is based on the quantile transformation. Consider more generally a stationary ergodic sequence $Y_1, Y_2, \ldots$ with marginal distribution function F which satisfies Assumption [C]. If $V_1, V_2, \ldots$ are independent $\mathcal{U}(0,1)$-distributed random variables, independent of the Y_i, then the sequence $U_n = F(Y_n-) + V_n\,\big(F(Y_n) - F(Y_n-)\big)$, $n \in \mathbb{N}$, is also stationary and ergodic with uniform marginals and $Y_n = F^{\mathrm{Inv}}(U_n)$ holds almost surely.

To apply Theorem 2.15 in the proof of Theorem 5.6, we let $\rho_n = G_n^{\mathrm{Inv}}$ be the quantile process based in $U_1, \ldots, U_n$, $\rho = G^{\mathrm{Inv}}$ be the quantile function of the uniform distribution, and set $\zeta(\varphi) = F^{\mathrm{Inv}}(\mathrm{id}_{[\alpha_l,\alpha_u]} + \varphi)$, which is semi-Hadamard differentiability with respect to the hypi-distance by Lemma 5.16 below. The distance in (2.5) will be zero almost surely. For the weak convergence in (2.7) with $a_n = \sqrt{n}$, or equivalently of $\sqrt{n}(G_n - G)$ for the uniform distribution function G and the empirical counterpart G_n, we may apply results for empirical processes for dependent data as presented, for example, in Oliveira and Suquet (1998), Dehling, Durieu, et al. (2009) or Dehling, Mikosch, et al. (2012). ◇

5.10 Example (Expectile process).
For proving the weak convergence of the expectile process, we let $\vartheta(\cdot) = \mu_\cdot$, $\vartheta_n(\cdot) = \widehat{\mu}_{\cdot,n}$, and $\xi_0 = \psi_0$ be defined by $[\psi_0(\varphi)](\tau) = -\mathbb{E}\left[I_\tau(\varphi(\tau); Y)\right]$ for $\tau \in [\tau_l, \tau_u]$ and $\varphi \in \ell^\infty([\tau_l, \tau_u])$. Additionally, we use $\rho_n = \xi_0(\vartheta_n)$ and $\rho = \xi_0(\rho)$; see the beginning of the proof of Theorem 5.1 in Section 5.6. The inverse functional $\zeta = \psi_0^{\mathrm{Inv}}$ is semi-Hadamard differentiable by Lemma 5.15. Further, Lemma 2.16 is applied with

$$[\xi_n(\varphi)](\tau) = -\frac{1}{n}\sum_{k=1}^{n} I_\tau(\varphi(\tau); Y_k)$$

and $a_n = \sqrt{n}$; see Lemma 5.13 below. If we estimate the expectile curve based on a stationary ergodic sequence $Y_1, Y_2, \ldots$ with marginal distribution function F, we need to verify the Donsker property (2.8) as well as (2.9), which is based on a maximal inequality in the case of independent identically distributed random variables. Corresponding results for empirical processes of strongly mixing sequences are given, for example, in Merlevède and Peligrad (2011), Rio (2013) and Andrews and Pollard (1994). ◇

In both examples we need the weak convergence of a transformed process and a (solely analytical) differentiability property. As we argued, the weak convergence of the transformed process can be generalized to other sequences of random variables, which is especially important in financial applications. On the other hand, the assumptions implying the semi-Hadamard differentiability cannot be weakened in the case for the expectile process, and only slightly for the quantile process; see Section 5.5.

5.3 Alternative proof for the quantile process

Here we discuss an alternative approach to Theorem 5.6. As we indicated in the former section, after some reductions it is enough to show semi-Hadamard differentiability of $F^{\mathrm{Inv}}(\mathrm{id}_{[\alpha_l,\alpha_u]} + \varphi)$. Another possibility is to show the following.

5.11 Theorem.

Consider the map $\Phi : (\mathcal{D}_0, \|\cdot\|) \longrightarrow (\ell^\infty([\alpha_l, \alpha_u]), \mathrm{d}_{hypi})$, $\Phi(h) = h^{\mathrm{Inv}}$ *for a subset* $\mathcal{D}_0 \subset \mathcal{D}([q_{\alpha_l}, q_{\alpha_u}])$ *where building the inverse* $(\cdot)^{\mathrm{Inv}}$ *is well-defined. Under Assumption [C] the map* Φ *is semi-Hadamard differentiable with respect to* d_{hypi} *in* F *tangentially to* $\mathbb{W} = \{\varphi \in \mathcal{D}([q_{\alpha_l}, q_{\alpha_u}]) \,|\, \varphi$ *jumps at most if* F *jumps*$\}$. *The semi-derivative is* $\dot{\Phi}(\varphi) = -\varphi \circ F^{\mathrm{Inv}}\, \partial^-(F^{\mathrm{Inv}})$.

With the functional delta-method we then directly deduce the following.

5.12 Corollary.

Let $Y_1, Y_2, \dots$ *be a stationary ergodic sequence with marginal distribution function* F *fulfilling Assumption [C]. Further, suppose that the process* $\sqrt{n}(F_n - F)$ *converges weakly in* $(\ell^\infty([q_{\alpha_l}, q_{\alpha_u}]), \|\cdot\|)$ *to a process* Z *which concentrates on* $\mathbb{W}$ *from Theorem 1. Then* $\sqrt{n}(F_n^{\mathrm{Inv}} - F^{\mathrm{Inv}})$ *converges weakly in distribution in* $(\ell^\infty([\alpha_l, \alpha_u]), \mathrm{d}_{hypi})$ *to* $\big(-\partial^-(F^{\mathrm{Inv}})\, Z \circ F^{\mathrm{Inv}}\big)$.

Note that this reduces to Theorem 5.6 if $Z = -V \circ F$ as

$$-\partial^-(F^{\mathrm{Inv}})(\alpha)\,(-V_{F(F^{\mathrm{Inv}}(\alpha))} = \begin{cases} \partial^-(F^{\mathrm{Inv}})(\alpha)\, V_\alpha & \text{if } F \text{ is continuous in } \alpha, \\ 0 & \text{else.} \end{cases}$$

The proof of Theorem 5.11, which is given in Section 5.6, is more involved than the one of Theorem 5.6 and so far does not include a possible Bootstrap-procedure. The latter is possible future research.

5.4 Numerical illustrations

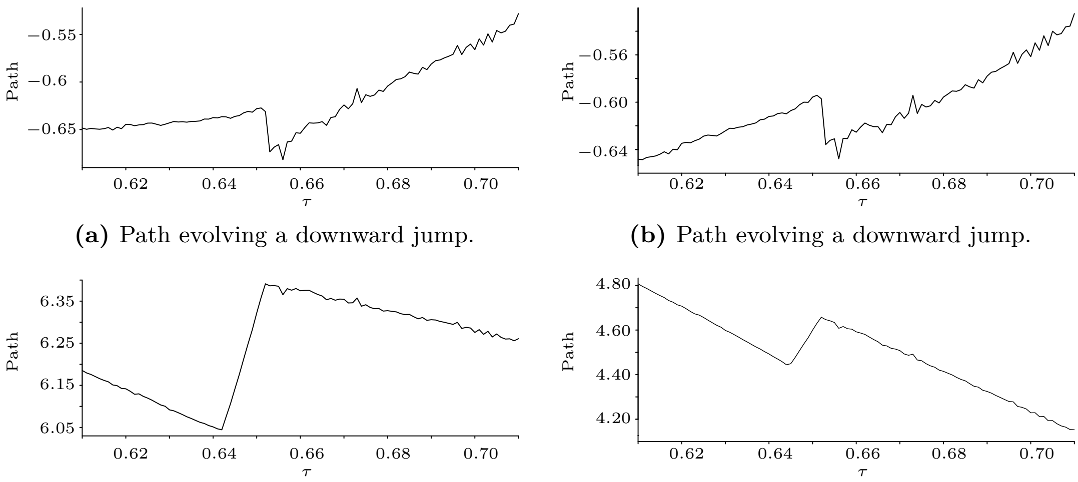
In this section we illustrate the asymptotic results for the expectile and quantile processes in a short simulation. For both processes we consider a random variable Y with distribution function given by

$$F(x) = \frac{9}{10} \int_{-\infty}^{x} \frac{1}{4\sqrt{2\pi}} \exp\Big(-\tfrac{y^2}{32}\Big)\, \mathrm{d}y + \frac{1}{10}\, \mathbb{1}(x \geq 1),$$

which is a mixture of a $\mathcal{N}(0,16)$ random variable and a point mass in 1, so it holds that $\mathbb{E}[Y] = 1/10$ and $\mathbb{E}[Y^2] = 14.5$. We visualize simulated paths both processes; thereafter we concentrate on the weak convergence of the supremum norm of the processes.

5.4.1 Illustrations for the expectile process

We first investigate results from Section 5.2.1 for the expectile process. Using equation (2.7) in Newey and Powell (1987), we numerically find $\mu_{\tau_0} = 1$ for $\tau_0 \approx 0.6529449$ and examine the expectile process on the interval $[0.6, 0.7]$.



(a) Path evolving a downward jump.

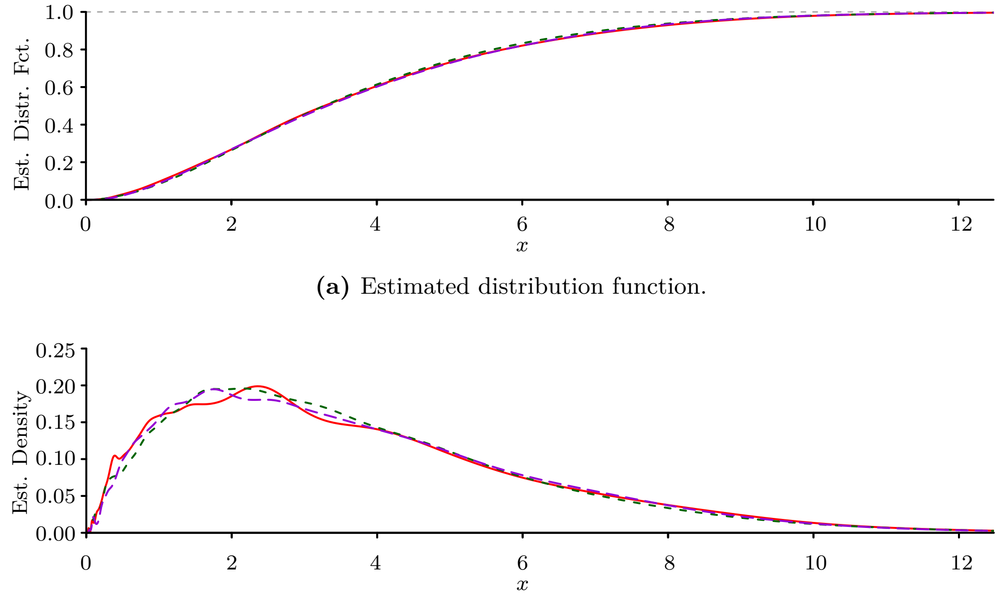
(b) Path evolving a downward jump.

(c) Path evolving an upward jump.

(d) Path evolving an upward jump.

Figure 5.1: The pictures show simulated paths of the empirical expectile process based on $n = 10^4$ observations of Y. If the path is negative (positive) around τ_0, a downward (upward) jump seems to evolve. This is plausible when considering the form of the lower- and upper-semicontinuous hulls of $\dot{\psi}^{\mathrm{Inv}}$ in the limit process.

In order to visualize the potential discontinuity of the paths of the empirical expectile process, we show four exemplary paths of $\sqrt{n}(\hat{\mu}_{\tau,n} - \mu_\tau)$ for samples of size $n = 10^4$ (Figure 5.1). All plotted paths seem to evolve a jump around τ_0, which is what we expect when considering the shape of the stated limit process in Theorem 5.1.



(a) Estimated distribution function.

(b) Estimated density.

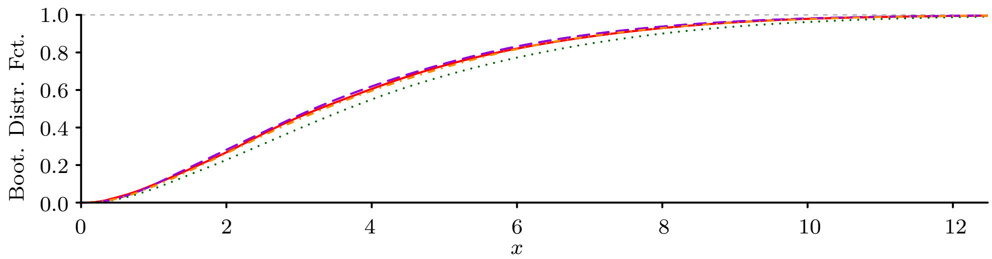
Figure 5.2: Figure (a) shows the cumulative distribution function of the supremum norm of $\sqrt{n}(\widehat{\mu}_{\cdot,n} - \mu_{\cdot})$, based on 10^4 samples of sizes $n = 10^2$ (green dashed), $n = 10^3$ (violet long-dashed) and $n = 10^4$ (red solid). Figure (b) shows the corresponding density estimates using the same colour code.

Table 5.1: Empirical quantiles for the supremum norm $\sqrt{n}\|\widehat{\mu}_{\cdot,n} - \mu_{\cdot}\|$, based on 10^4 samples of sizes $n \in \{10^2, 10^3, 10^4\}$ and last averaged over $2 \cdot 10^2$ repetitions. The terms in brackets are the resulting standard deviations.

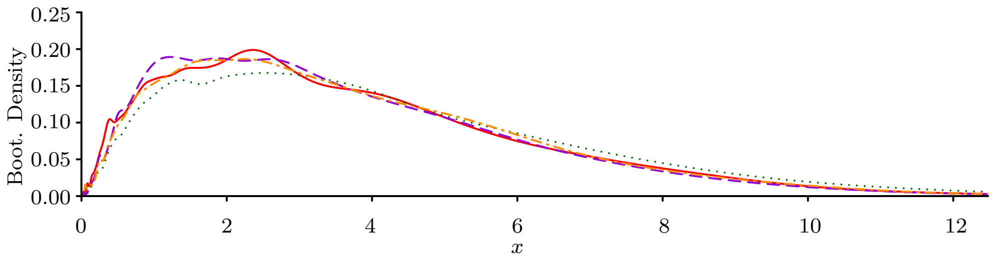
Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	0.338 (0.015)	0.740 (0.017)	1.076 (0.017)	1.900 (0.023)	3.313 (0.033)	5.195 (0.038)	7.155 (0.047)	8.374 (0.069)	10.744 (0.135)
10^3	0.336 (0.015)	0.736 (0.018)	1.072 (0.020)	1.895 (0.023)	3.307 (0.030)	5.200 (0.043)	7.168 (0.054)	8.425 (0.072)	10.880 (0.14)
10^4	0.339 (0.015)	0.740 (0.017)	1.074 (0.019)	1.897 (0.023)	3.305 (0.030)	5.182 (0.038)	7.137 (0.061)	8.385 (0.074)	10.833 (0.149)

Next, we investigate the distribution of the supremum norm of the expectile process on the interval $[0.6, 0.7]$. To this end, we simulate 10^4 samples of sizes $n \in \{10^2, 10^3, 10^4\}$, compute the expectile process and determine its supremum norm. This is repeated $2 \cdot 10^2$ times. Exemplary plots of the resulting empirical distribution function and density estimate of this statistic are contained in Figure 5.2. Additionally, we report (averaged) quantiles of the estimated distribution functions at various levels, which support the visual impression of quick convergence of the distribution of the supremum norm (Table 5.1).

Finally, to illustrate performance of the n-out-of-n bootstrap, we display exemplary bootstrap distribution functions obtained from 10^4 bootstrap samples of $\sqrt{n}\|(\mu^*_{\cdot,n} - \hat{\mu}_{\cdot,n})\|$ based on a single sample of size $n \in \{10^2, 10^3, 10^4\}$, together with the distribution of $\sqrt{n}\|(\hat{\mu}_{\cdot,n} - \mu_{\cdot})\|$ with $n = 10^4$ as a reference (Figure 5.3).



(a) Estimated bootstrap distribution function.



(b) Estimated bootstrap density.

Figure 5.3: Figure (a) shows the estimated cumulative bootstrap distribution function of $\sqrt{n}\|(\mu^*_{\cdot,n} - \widehat{\mu}_{\cdot,n})\|$ estimated from 10^4 samples of an underlying sample of size $n = 10^2$ (green dashed), $n = 10^3$ (violet long dashed) and $n = 10^4$ (orange dot-dashed). Figure (b) contains the estimated density thereof with the same colour code. The solid red line indicates the estimated empirical distribution and density function, respectively, as in Figure 5.2.

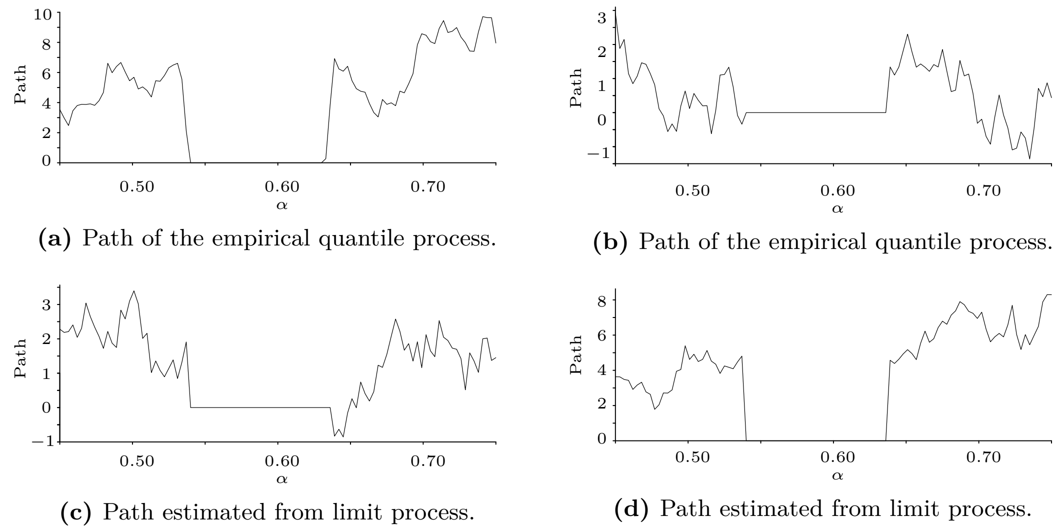
We moreover report the quantiles of the estimated bootstrap distribution functions averaged over $2 \cdot 10^2$ repetitions of the simulation, together with the respective standard deviations (Table 5.2). The bootstrap distribution function for $n = 10^4$ is quite close to the empirical one, which is also valid for the bootstrap quantiles. Additionally, the latter are rather stable.

Table 5.2: Bootstrap quantiles for the supremum norm $\sqrt{n}\|\mu^*_{\cdot,n} - \widehat{\mu}_{\cdot,n}\|$, obtained from 10^4 estimates of this statistic, averaged over $2 \cdot 10^2$ repetitions. The bracketed numbers are the calculated standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^2	0.339	0.734	1.065	1.885	3.300	5.190	7.155	8.406	10.842
	(0.032)	(0.062)	(0.088)	(0.152)	(0.269)	(0.431)	(0.606)	(0.714)	(0.938)
10^3	0.344	0.741	1.074	1.901	3.321	5.215	7.195	8.465	10.938
	(0.016)	(0.025)	(0.031)	(0.051)	(0.092)	(0.143)	(0.202)	(0.250)	(0.355)
10^4	0.345	0.742	1.077	1.904	3.324	5.217	7.200	8.469	10.955
	(0.015)	(0.019)	(0.020)	(0.027)	(0.043)	(0.064)	(0.087)	(0.108)	(0.162)
Emp. 10^4	0.339	0.740	1.074	1.897	3.305	5.182	7.137	8.385	10.833

5.4.2 Illustrations for the quantile process

Next, we consider the results given in Section 5.2.2 for the same distribution function F as before. Here, we are interested in the behaviour of the quantile process in the interval $[\alpha_l, \alpha_u]$, such that $q_\alpha = 1$ for every α in some subinterval of $[\alpha_l, \alpha_u]$. This subinterval is given by $[F(1-), F(1)]$, with $F(1-) \approx 0.5388$ and $F(1) \approx 0.6388$; to capture the behaviour in this set we consider the quantile process in the interval [0.45, 0.75].



(a) Path of the empirical quantile process.

(b) Path of the empirical quantile process.

(c) Path estimated from limit process.

(d) Path estimated from limit process.

Figure 5.4: The images show simulated paths of the (empirical) quantile process. The above two images (a) and (b) were obtained from $n = 10^4$ observations of Y and calculating the respective empirical quantile process. The below two images (c) and (d) are simulated directly from the asserted limit process given in Theorem 5.6.

Note that F is differentiable in $y \in \mathbb{R} \setminus \{1\}$ with derivative given by

$$F'(y) = \tfrac{9}{40\sqrt{2\pi}} \exp\left(\tfrac{y^2}{32}\right);$$

in $y = 1$ it holds that

$$\partial^+(F)(1) = \tfrac{9}{40\sqrt{2\pi}} \exp\left(\tfrac{1}{32}\right) = \partial^-(F)(1-).$$

In addition, $\partial^-(F)(1) = \infty$ is true, such that F fulfils Assumption [C] and Theorem 5.6 implies

$$\sqrt{n}(\widehat{q}_{\alpha,n} - q_\alpha) \rightsquigarrow F'(q_\alpha)^{-1} V_\alpha \mathbb{1}(\alpha \notin [F(1-), F(1)])$$

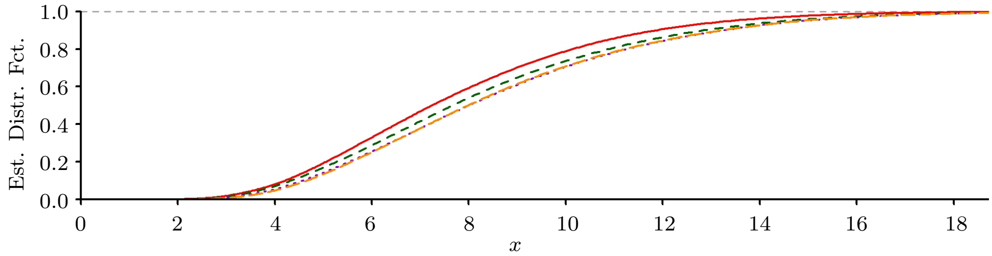
in $\left(\ell^\infty([0.45,\ 0.75]), \mathrm{d}_{hypi}\right)$ with a standard Brownian Bridge $(V_\alpha)_{\alpha\in[0,1]}$. Hence, we expect the empirical quantile process to be almost zero around the interval $[F(1-), F(1)]$. This behaviour is visualized in Figure 5.4. There we first show simulated paths of the empirical quantile process (Figures 5.4 (a) and (b)) obtained from an independent identically distributed sample of size $n = 10^4$ from Y; second, we directly simulate paths from the asserted limit process by transforming simulated paths of a standard Brownian Bridge with $\partial^\pm(F^{\mathrm{Inv}})$ (Figures 5.4 (c) and (d)). In this case we used $n = 3 \cdot 10^4 + 1$ sampling points which were equally distributed in $[0.45, 0.75]$.

Now, we investigate the distribution of the supremum norm of the quantile process restricted to the interval $[0.45,\ 0.75]$. Therefore we simulated 10^4 samples from Y of size $n \in \{10^3, 5 \cdot 10^4, 10^5\}$, computed the empirical quantile process and calculated the supremum thereof. This was iterated $2 \cdot 10^2$ times. An exemplary empirical distribution function and density estimate of the statistic can be found in Figure 5.5. These estimates are compared to the theoretical limit, whose distribution function is approximated by simulating 10^4 paths of $\left(\partial^\pm(F^{\mathrm{Inv}})\, V\right)$, computing the supremum norm thereof and finally estimating the distribution function. In addition, we report the (averaged) empirical quantiles at various levels in comparison to the quantiles obtained directly by sampling from the theoretical limit (Table 5.3).

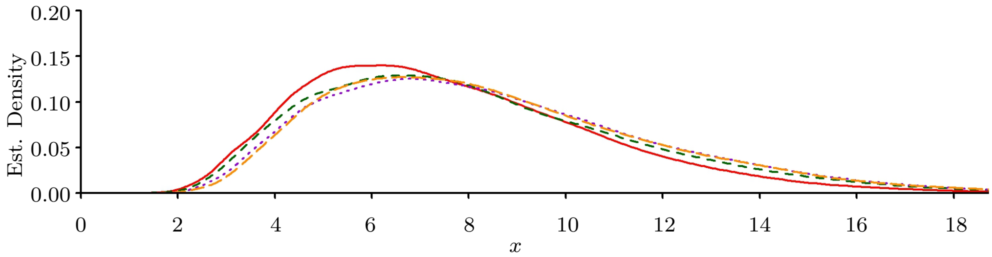
5.5 Conclusion and discussion

We showed the weak convergence of the empirical expectile and quantile processes and derived statistics under weak assumptions on the underlying distribution function. We also obtained bootstrap consistency in both cases.

We were able to renounce the standard assumptions of differentiability of the distribution function in the quantile case to some extent. Considering the assumptions on the possible point masses of F in Assumption [C], we conjecture that this can still be weakened, as F can possibly admit countably many point masses $\{y_i\}_{i\in\mathbb{N}}$, such that $\{y_i\}$ has at most finitely many accumulation points. In the corresponding levels $F(y_i-)$ and



(a) Estimated distribution function.



(b) Estimated density.

Figure 5.5: Figure (a) shows the estimated cumulative distribution function of $\sqrt{n}\|\widehat{q}_{\cdot,n} - q_{\cdot}\|$ estimated from 10^4 samples of size $n = 10^3$ (green dashed), $n = 5 \cdot 10^4$ (violet dotted) and $n = 10^5$ (orange long-dashed). Figure (b) contains the estimated density thereof with the same colour code. The solid red lines indicate the respective functions estimated directly from the theoretical limit.

Table 5.3: Empirical quantiles for the supremum norm $\sqrt{n}\|\widehat{q}_{\cdot,n} - q_{\cdot}\|$, based on 10^4 samples of sizes $n \in \{10^3, 5 \cdot 10^4, 10^5\}$ and last averaged over $2 \cdot 10^2$ repetitions. The last line is simulated directly from the theoretical limit. The terms in brackets indicate the calculated standard deviations.

Size	Quantile								
n	1%	5%	10%	25%	50%	75%	90%	95%	99%
10^3	2.896 (0.041)	3.776 (0.034)	4.394 (0.034)	5.723 (0.037)	7.694 (0.041)	10.171 (0.051)	12.728 (0.071)	14.355 (0.093)	17.563 (0.182)
$5 \cdot 10^4$	3.124 (0.041)	4.019 (0.030)	4.647 (0.032)	5.991 (0.033)	7.98 (0.045)	10.485 (0.058)	13.076 (0.079)	14.723 (0.095)	17.959 (0.181)
10^5	3.134 (0.040)	4.027 (0.034)	4.652 (0.037)	6.000 (0.036)	7.995 (0.039)	10.506 (0.054)	13.094 (0.079)	14.749 (0.104)	17.995 (0.189)
True	2.825 (0.041)	3.636 (0.030)	4.202 (0.031)	5.407 (0.035)	7.202 (0.044)	9.461 (0.048)	11.804 (0.069)	13.297 (0.088)	16.234 (0.153)

$F(y_i)$, the right- and left-sided derivatives $\partial^-(F)$ and $\partial^+(F)$ must exist but need not be càdlàg or làdcàg. Under these assertions we should still be able to determine the lower- and upper-semicontinuous hulls of $\partial^-(F^{\mathrm{Inv}})$ and construct sequences attaining the values of $(\partial^-(F^{\mathrm{Inv}}))_\vee$ or $(\partial^-(F^{\mathrm{Inv}}))_\wedge$ in $F(y_i-)$ and $F(y_i)$.

On the other hand, after first considerations, it is questionable whether it is possible to further work toward Assumption [A] in Chapter 3, especially to obtain a slower rate of convergence for the empirical quantile process.

For expectiles we needed $Y \in \mathcal{L}_2$ in order to obtain a finite limiting variance, but nothing beyond. Dropping the assumption of finite second moments for the empirical expectiles leads to stable limit distribution for individual expectiles; see Holzmann and Klar (2016). Possibly this result can be generalized to process convergence.

5.6 Proofs

Here we present the proofs of the assertions in the current chapter. First, we give outlines of the proofs to highlight the ideas. The details are exposed at the end of this section, whereas some technical calculations are further deferred to Section 5.9.

5.6.1 Outline of the proofs of main results

We sketch how to prove the main theorems on the convergence of the empirical and bootstrap expectile process, followed by a scheme for the proofs of the respective assertions for quantiles.

5.6.1.1 Outline of the proofs of Theorems 5.1 and 5.4

The proofs of Theorems 5.1 and 5.4 are both separated into several steps, which aim for an application of the functional delta-method.

Proof of Theorem 5.1 (Outline). For a distribution function $Q \in \mathcal{L}_1$ let us define $\psi(\varphi; Q)(\tau) = -I_\tau(\varphi(\tau); Q) = -\int I_\tau(\varphi(\tau); y)\,\mathrm{d}Q(y)$, where $\tau \in [\tau_l, \tau_u]$ and $\varphi \in \ell^\infty([\tau_l, \tau_u])$, and set $\psi_0(\cdot) = \psi(\cdot; F)$ and $\psi_n(\cdot) = \psi(\cdot; F_n)$.

We proceed by proving the following steps.

Step 1 Weak convergence of $\sqrt{n}(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot))$ to Z in $(\ell^\infty([\tau_l, \tau_u]), \|\cdot\|)$.

Step 2 Invertibility of ψ_0 and semi-Hadamard differentiability of the inverse with respect to d_{hypi}.

Step 3 Conclusion with the generalized functional delta-method.

These parts actually show that the assertions of Lemma 2.16 and Theorem 2.15 are fulfilled as already hinted at in Section 5.2.3.

Step 1 Weak convergence of $\sqrt{n}\big(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot)\big)$ to Z in $\big(\ell^\infty([\tau_l,\tau_u]), \|\cdot\|\big)$.

This step shows (2.6) for the case of expectiles and uses standard results from empirical process theory based on bracketing properties of Lipschitz-continuous functions. The main issue in the proof of the following lemma is the Lipschitz-continuity of $\tau \mapsto \mu_\tau$, $\tau \in [\tau_l, \tau_u]$, for a general distribution function F, which is of some interest in itself; see Lemma 5.19.

5.13 Lemma.

In $(\ell^\infty([\tau_l,\tau_u]), \|\cdot\|)$ we have the weak convergence

$$\sqrt{n}\big(\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot)\big)(\tau) \rightsquigarrow Z_\tau, \quad \tau \in [\tau_l, \tau_u]. \tag{5.2}$$

Further, given $\delta_n \searrow 0$ we have as $n \to \infty$ that

$$\sup_{\|\varphi\| \le \delta_n} \sqrt{n}\,\big\|\psi_n(\mu_\cdot + \varphi)(\cdot) - \psi_0(\mu_\cdot + \varphi)(\cdot) - \big[\psi_n(\mu_\cdot)(\cdot) - \psi_0(\mu_\cdot)(\cdot)\big]\big\| = o_{\mathrm{P}}(1). \tag{5.3}$$

The assertions of Lemma 5.13 are the translations of (2.8) and (2.9).
Since $\psi_0(\mu_\cdot) = \psi_n(\hat{\mu}_{\cdot,n}) = 0$, we can rewrite

$$\begin{aligned}\sqrt{n}\big(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot)\big) &= \sqrt{n}\big(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_n(\widehat{\mu}_{\cdot,n})\big)\\ &= \sqrt{n}\big[\psi_0(\mu_\cdot + \varphi_n) - \psi_n(\mu_\cdot + \varphi_n)\big]\end{aligned} \tag{5.4}$$

for $\varphi_n(\cdot) = \widehat{\mu}_{\cdot,n} - \mu_\cdot$ and adding and subtracting $-\sqrt{n}\big(\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot)\big)$ yields

$$\begin{aligned}&\sqrt{n}\big(\psi_0(\mu_\cdot + \varphi_n(\cdot)) - \psi_n(\mu_\cdot + \varphi_n(\cdot))\big)\\ &= -\sqrt{n}\big(\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot)\big) + \sqrt{n}\Big[\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot) - \big(\psi_n(\mu_\cdot + \varphi_n) - \psi_0(\mu_\cdot + \varphi_n)\big)\Big].\end{aligned} \tag{5.5}$$

Due to the uniform consistency shown in Theorem 2, Holzmann and Klar (2016), it holds that $\|\varphi_n\| = o_{\mathrm{P}}(1)$, such that the supremum (over $\tau \in [\tau_l, \tau_u]$) of the term in angle brackets above is smaller than (or equal to) the expression in (5.3). Using this together with (5.4) and (5.5) shows

$$\sqrt{n}\big(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot)\big) = -\sqrt{n}\big(\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot)\big) + o_{\mathrm{P}}(1).$$

Then (5.2) and the fact that Z and $-Z$ have the same law conclude the proof of

$$\sqrt{n}\big(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot)\big) \rightsquigarrow Z \tag{5.6}$$

in $\big(\ell^\infty([\tau_l,\tau_u]), \|\cdot\|\big)$, finishing Step 1.

Step 2 Invertibility of ψ_0 and semi-Hadamard differentiability of the inverse with respect to d_{hypi}.

This step shows that $\zeta = \psi_0^{\mathrm{Inv}}$ can be defined and fulfils the assertion of Theorem 2.15. The first part of this step is made precise in the next lemma.

5.14 Lemma.

The map ψ_0 is invertible. Further, $\psi_0^{\mathrm{Inv}}(\varphi) \in \ell^\infty[\tau_l, \tau_u]$ for any $\varphi \in \ell^\infty[\tau_l, \tau_u]$.

The next result then is the key technical ingredient in the proof of Theorem 5.1.

5.15 Lemma.

The map ψ_0^{Inv} is semi-Hadamard differentiable with respect to the hypi-semimetric in $0 \in \mathcal{C}([\tau_l, \tau_u])$ tangentially to $\mathcal{C}([\tau_l, \tau_u])$ with semi-Hadamard derivative given by $\dot{\psi}_0^{\mathrm{Inv}}(\varphi)(\tau) = \big(\tau + (1 - 2\,\tau)F(\mu_\tau)\big)^{-1}\varphi(\tau)$, $\varphi \in \ell^\infty([\tau_l, \tau_u])$, that is, we have

$$t_n^{-1}\big(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n) - \psi_0^{\mathrm{Inv}}(0)\big) \to \dot{\psi}_0^{\mathrm{Inv}}(\varphi)$$

for any sequence $t_n \to 0$, $t_n > 0$ and $\varphi_n \in \ell^\infty([\tau_l, \tau_u])$, where $\varphi_n \to \varphi \in \mathcal{C}([\tau_l, \tau_u])$ with respect to d_{hypi}.

The proof of the lemma is based on an explicit representation of increments of ψ_0^{Inv}, and Lemmas 2.9 and 2.13.

We observe here that up to the former lemma the results did not depend on the hypi-semimetric. When wanting to obtain Theorem 5.1 directly for the M_2-topology, we can use the same steps, just replacing hypi- with M_2-convergence in the above lemma. This is where Lemma 2.23 steps in, substituting the application of Lemma 2.13. We will give more details for this in Section 5.8.

Step 3 Conclusion with the generalized functional delta-method.

Due to the steps 1 and 2 we can conclude with Theorem 2.15. Precisely, from (5.6), Lemma 5.15 and the generalized functional delta-method, Theorem B.7, Bücher et al. (2014), we obtain

$$\sqrt{n}(\widehat{\mu}_{\cdot,n} - \mu_\cdot) = \sqrt{n}\Big(\psi_0^{\mathrm{Inv}}(\psi_0(\widehat{\mu}_{\cdot,n})) - \psi_0^{\mathrm{Inv}}(0)\Big) \rightsquigarrow \dot{\psi}_0^{\mathrm{Inv}}(Z)$$

in $(\ell^\infty([\tau_l, \tau_u]), \mathrm{d}_{hypi})$. □

Note that the generalized functional delta-method is formulated for arbitrary (semi-)

metric spaces, such that the conclusion would also work with respect to M_2-convergence, given Lemma 5.15 is proven with respect to the M_2-topology as well.

For the bootstrap version we proceed similar as in the proof of Theorem 5.1.

Proof of Theorem 5.4 (Outline). In the analogous result to Lemma 5.13 and (5.6) (see Step 1 above), we require the almost sure uniform consistency of $\mu^*_{\cdot,n}$ as in Holzmann and Klar (2016), Theorem 2. The weak convergence statements in that step require the changing classes central limit theorem, van der Vaart (1998), Theorem 19.28. In the second step we argue directly with the extended continuous mapping theorem, Theorem B.3 in Bücher et al. (2014). □

5.6.1.2 Outline of the proofs of Theorem 5.6 and 5.7

Here we sketch the proof of Theorem 5.6 and already prove Theorem 5.7 in detail; both proofs depend on the quantile transformation. The proof of Lemma 5.5 is further relegated to Section 5.6.2.3.

Again we drop the subscript in $\|\cdot\|_{[\alpha_l,\alpha_u]}$ and just write $\|\cdot\|$.

Proof of Theorem 5.6 (Outline). Let G be the distribution function of $\mathcal{U}(0,1)$ and let G_n be the empirical distribution function of an independent, $\mathcal{U}(0,1)$-distributed sample $U_1,\ldots,U_n$. By the quantile transformation we can write

$$\sqrt{n}(\widehat{q}_{\cdot,n} - q_\cdot) = \sqrt{n}\big(F^{\mathrm{Inv}}(G_n^{\mathrm{Inv}}(\cdot)) - F^{\mathrm{Inv}}(G^{\mathrm{Inv}}(\cdot))\big).$$

The process $\sqrt{n}(G_n^{\mathrm{Inv}} - G^{\mathrm{Inv}})$ converges in distribution in $(\ell^\infty([\alpha_l,\alpha_u]),\|\cdot\|)$ to a Brownian bridge V; see Example 21.6, van der Vaart (1998).

To use a functional delta-method for the hypi-semimetric, we require

$$t_n^{-1}\big(F^{\mathrm{Inv}}(\alpha + t_n\,\varphi_n(\alpha)) - F^{\mathrm{Inv}}(\alpha)\big) \longrightarrow \partial^-(F^{\mathrm{Inv}})(\alpha)\,\varphi(\alpha)$$

with respect to the hypi-distance for $t_n \to 0$, $t_n > 0$ and $\varphi_n \in \ell^\infty([\alpha_l,\alpha_u])$, $\varphi \in \mathcal{C}([\alpha_l,\alpha_u])$ such that $\mathrm{d}_{hypi}(\varphi_n,\varphi) \to 0$. That is the semi-Hadamard differentiability of the functional $\Psi_0^{\mathrm{Inv}}(\nu)(\alpha) = F^{\mathrm{Inv}}(\alpha + \nu(\alpha))$, $\nu \in \ell^\infty([\alpha_l,\alpha_u])$ with respect to the hypi-distance as in Definition 2.14. To also be able to deal with the bootstrap version, we directly show a slightly stronger version, the *uniform* semi-Hadamard differentiability.

5.16 Lemma.

Let $t_n \to 0$, $t_n > 0$, $\varphi_n,\nu_n \in \ell^\infty([\alpha_l,\alpha_u])$ and $\varphi \in \mathcal{C}([\alpha_l,\alpha_u])$ with $\mathrm{d}_{hypi}(\varphi_n,\varphi) \to 0$ and $\mathrm{d}_{hypi}(\nu_n,0) \to 0$. Under Assumption [C], we have the hypi-convergence

$$t_n^{-1}\Big[F^{\mathrm{Inv}}(\mathrm{id}_{[\alpha_l,\alpha_u]}(\cdot)+\nu_n(\cdot)+t_n\,\varphi_n(\cdot)) - F^{\mathrm{Inv}}(\mathrm{id}_{[\alpha_l,\alpha_u]}(\cdot)+\nu_n(\cdot))\Big] \longrightarrow \partial^-(F^{\mathrm{Inv}})(\cdot)\,\varphi(\cdot).$$

Using the convergence of $\sqrt{n}(G_n^{\mathrm{Inv}} - G^{\mathrm{Inv}})$, Lemma 5.16 and the functional delta-method, Theorem B.7, Bücher et al. (2014), we conclude

$$\sqrt{n}(\widehat{q}_{\cdot,n} - q_\cdot) = \sqrt{n}\Big[F^{\mathrm{Inv}}(G_n^{\mathrm{Inv}}(\cdot)) - F^{\mathrm{Inv}}(G^{\mathrm{Inv}}(\cdot))\Big] \rightsquigarrow \partial^-(F^{\mathrm{Inv}})(\cdot)\, V_\cdot \tag{5.7}$$

in $(\ell^\infty([\alpha_l, \alpha_u]), \mathrm{d}_{hypi})$. □

The next proof depends on Bücher and Kojadinovic (2018), who give a connection between conditional and unconditional weak convergence in bootstrap consistency results.

Proof of Theorem 5.7. Given a sample $U_1, \ldots, U_n$ of independent $\mathcal{U}(0,1)$-distributed random variables and the corresponding empirical distribution function G_n we denote with $U_1^{[i]}, \ldots, U_n^{[i]}$, $i = 1, 2$, two independent bootstrap samples – that is drawn with replacement from $U_1, \ldots, U_n$ – with bootstrap distribution functions $G_n^{[i]}$. We set the bootstrap quantiles $q_{\cdot,n}^{[i]} = q_\cdot(G_n^{[i]})$ and shall show that

$$\Big(\sqrt{n}(\widehat{q}_{\cdot,n} - q_\cdot),\, \sqrt{n}(q_{\cdot,n}^{[1]} - \widehat{q}_{\cdot,n}),\, \sqrt{n}(q_{\cdot,n}^{[2]} - \widehat{q}_{\cdot,n})\Big) \rightsquigarrow H(\cdot)(V_\cdot, V_\cdot^{[1]}, V_\cdot^{[2]}) \tag{5.8}$$

weakly in $(\ell^\infty([\alpha_l, \alpha_u]), d_{hypi})^3$, where $H(\alpha) = \partial^-(F^{\mathrm{Inv}})(\alpha)$ for brevity and $V^{[1]}$ and $V^{[2]}$ are independent copies of V. Then weak consistency of the bootstrap, that is, $\sqrt{n}(q_{\cdot,n}^{[1]} - \widehat{q}_{\cdot,n}) \rightsquigarrow H(\cdot)\, V_\cdot^{[1]}$ conditionally on $Y_1, Y_2, \ldots$ in probability in $\big(\ell^\infty([\alpha_l, \alpha_u]), d_{hypi}\big)$ follows from Lemma 2.2 in Bücher and Kojadinovic (2018).

In order to show (5.8) we observe that by Example 3.9.24, van der Vaart and Wellner (1996), the process $\sqrt{n}(G_n^{\mathrm{Inv}} - G^{\mathrm{Inv}})$ converges in distribution to V in $\big(\ell^\infty([\alpha_l, \alpha_u]), \|\cdot\|\big)$; using Theorem 3.9.11 in combination with Lemma 3.9.23 in that reference yields

$$\sqrt{n}((G_n^{[i]})^{\mathrm{Inv}} - G_n^{\mathrm{Inv}}) \rightsquigarrow V^{[i]}$$

conditionally on $Y_1, Y_2, \ldots$ in probability in $\big(\ell^\infty([\alpha_l, \alpha_u]), \|\cdot\|\big)$. Then, using Corollary 2.9.3, van der Vaart and Wellner (1996), together with (5.7) we deduce that

$$\Big(\sqrt{n}(G_n^{\mathrm{Inv}} - G^{\mathrm{Inv}}),\, \sqrt{n}((G_n^{[1]})^{\mathrm{Inv}} - G_n^{\mathrm{Inv}}),\, \sqrt{n}((G_n^{[2]})^{\mathrm{Inv}} - G_n^{\mathrm{Inv}})\Big) \rightsquigarrow (V,\, V^{[1]}, V^{[2]})$$

in $\big(\ell^\infty([0,1]), \|\cdot\|\big)^3$. Using the quantile transformation we obtain

$$\sqrt{n}(q_{\cdot,n}^{[i]} - \widehat{q}_{\cdot,n}) = \sqrt{n}\Big[F^{\mathrm{Inv}}((G_n^{[i]})^{\mathrm{Inv}}(\cdot)) - F^{\mathrm{Inv}}(G_n^{\mathrm{Inv}}(\cdot))\Big] \tag{5.9}$$

and similar for $\sqrt{n}(\widehat{q}_{\cdot,n} - q_\cdot)$ (see (5.7)). Next, observe that we can write

$$(G_n^{[i]})^{\mathrm{Inv}} = G^{\mathrm{Inv}} + (G_n^{\mathrm{Inv}} - G^{\mathrm{Inv}}) + \frac{1}{\sqrt{n}}\sqrt{n}((G_n^{[i]})^{\mathrm{Inv}} - G_n^{\mathrm{Inv}}), \tag{5.10}$$

where $\|G_n^{\text{Inv}} - G^{\text{Inv}}\| \to 0$ almost surely and $\sqrt{n}((G_n^{[i]})^{\text{Inv}} - G_n^{\text{Inv}})$ converges weakly in the space $(\ell^\infty([\alpha_l, \alpha_u]), \|\cdot\|)$ to $V^{[i]}$ as shown above. Setting $t_n = n^{-1/2}$, $G^{\text{Inv}} = \mathrm{id}_{[\alpha_l,\alpha_u]}$, $\nu_n = G_n^{\text{Inv}} - G^{\text{Inv}}$ and $\varphi_n^{[i]} = \sqrt{n}((G_n^{[i]})^{\text{Inv}} - G_n^{\text{Inv}})$, (5.9) with (5.10) plugged in reads as

$$t_n^{-1}\Big[F^{\text{Inv}}\big(\mathrm{id}_{[\alpha_l,\alpha_u]} + \nu_n + t_n\,\varphi_n^{[i]}\big) - F^{\text{Inv}}\big(\mathrm{id}_{[\alpha_l,\alpha_u]} + \nu_n\big)\Big].$$

Here we see that we actually require the uniform semi-Hadamard differentiability of F^{Inv} from Lemma 5.16, as we have to deal with the additional variation between G_n^{Inv} and G^{Inv} coded in ν_n. Using this lemma in combination with the extended continuous mapping theorem, Corollary B.5, Bücher et al. (2014), concludes the proof of (5.8). □

5.6.2 Proofs of important parts

In the following we present the details for the steps needed in the proof of Theorem 5.1 and Theorem 5.4. We also prove the remaining part for concluding Theorem 5.6, namely Lemma 5.16.

5.6.2.1 Details for the proof of Theorem 5.1

Recall from Holzmann and Klar (2016) the identity

$$I_\tau(x;F) = \tau \int_x^\infty \big(1 - F(y)\big)\,\mathrm{d}y - (1-\tau)\int_{-\infty}^x F(y)\,\mathrm{d}y. \tag{5.11}$$

We start with stating some technical preliminaries. First, we are concerned with increments of the identification function I_τ.

5.17 Lemma.

We have that for $x_1, x_2 \in \mathbb{R}$,

$$I_\tau(x_1;F) - I_\tau(x_2;F) = (x_2 - x_1)\Big[\tau + (1-2\,\tau)\int_0^1 F\big(x_2 + s(x_1 - x_2)\big)\,\mathrm{d}s\Big]. \tag{5.12}$$

Now we state a bound which will guarantee that the inverse of ψ_0 still maps into $\ell^\infty([\tau_l, \tau_u])$.

5.18 Lemma.

For all $\tau \in [\tau_l, \tau_u]$, $s \in [0,1]$ it holds that

$$\min\big\{\tau_l, 1-\tau_u\big\} \le \tau + (1-2\tau)s \le 3/2. \tag{5.13}$$

Next, we discuss Lipschitz-properties of relevant maps.

5.19 Lemma.

For any $x_1, x_2, z \in \mathbb{R}$ and $\tau \in [\tau_l, \tau_u]$,

$$\big|I_\tau(x_1; z) - I_\tau(x_2; z)\big| \le |x_2 - x_1|. \tag{5.14}$$

Further, for any $\tau, \tau' \in [\tau_l, \tau_u]$ and $x, z \in \mathbb{R}$,

$$\big|I_\tau(x; z) - I_{\tau'}(x; z)\big| \le |\tau - \tau'| \,(|x| + |z|). \tag{5.15}$$

Finally, the map $\tau \mapsto \mu_\tau$, $\tau \in [\tau_l, \tau_u]$, is Lipschitz-continuous.

The proofs of Lemmas 5.17, 5.18 and 5.19 are given in Section 5.9.

We continue with proving the weak convergence of the process $\sqrt{n}(\psi_0(\widehat{\mu}_{\cdot,n}) - \psi_0(\mu_\cdot))$ comprised in Step 1.

Proof of Lemma 5.13. We start with the proof of (5.2). By Lemma 5.19 the class

$$\mathcal{H} = \big\{z \mapsto -I_\tau(\mu_\tau; z) \;\mid\; \tau \in [\tau_l, \tau_u]\big\}$$

consists of functions which are Lipschitz-continuous in the parameter τ for given z, where the Lipschitz constant (which depends on z) is square-integrable under F. Indeed, the triangle inequality first gives

$$\big|I_\tau(\mu_\tau; z) - I_{\tau'}(\mu_{\tau'}; z)\big| \le \big|I_\tau(\mu_\tau; z) - I_{\tau'}(\mu_\tau; z)\big| + \big|I_{\tau'}(\mu_\tau; z) - I_{\tau'}(\mu_{\tau'}; z)\big|.$$

Using (5.15) the first addend on the right hand side fulfils

$$\big|I_\tau(\mu_\tau; z) - I_{\tau'}(\mu_\tau; z)\big| \le |\tau - \tau'| \,(|\mu_{\tau_l}| \vee |\mu_{\tau_u}| + |z|),$$

and the second is bounded by

$$\big|I_{\tau'}(\mu_\tau; z) - I_{\tau'}(\mu_{\tau'}; z)\big| \le |\mu_\tau - \mu_{\tau'}| \;\le |\tau - \tau'| \,\frac{|\mu_{\tau_u}| \vee |\mu_{\tau_l}| + \mathbb{E}\,[|Y|]}{\min\big\{\tau_l, 1 - \tau_u\big\}},$$

utilizing (5.14) and (5.40) below. Thus,

$$\big|I_\tau(\mu_\tau; z) - I_{\tau'}(\mu_{\tau'}; z)\big| \le |\tau - \tau'| \,(C + |z|)$$

is valid for some constant $C \ge 1$. By Example 19.7 in combination with Theorem 19.5 in van der Vaart (1998), $\mathcal{H}$ is a Donsker class, so that $\sqrt{n}(\psi_n(\mu_\cdot) - \psi_0(\mu_\cdot))$ converges weakly to the process Z. The same reasoning as in Theorem 8, Holzmann and Klar

(2016), then shows continuity of the sample paths of Z with respect to the Euclidean distance on $[\tau_l, \tau_u]$ as asserted.

Next, we prove (5.3). Setting

$$\mathcal{H}_{\delta_n} = \{z \mapsto I_\tau(\mu_\tau + x; z) - I_\tau(\mu_\tau; z) \mid |x| \leq \delta_n,\, \tau \in [\tau_l, \tau_u]\}$$

we estimate that

$$\sup_{\|\varphi\| \leq \delta_n} \sqrt{n} \left\| \psi_n(\mu_\cdot + \varphi)(\cdot) - \psi_0(\mu_\cdot + \varphi)(\cdot) - [\psi_n(\mu_\cdot)(\cdot) - \psi_0(\mu_\cdot)(\cdot)] \right\|$$

is smaller than (or equal to) $\|\mathbb{G}_n\|_{\mathcal{H}_{\delta_n}}$. We show convergence in probability of the latter to zero by utilising a maximal inequality depending on the bracketing integral. From the triangle inequality, for any $\tau, \tau' \in [\tau_l, \tau_u]$ and $x, x' \in [-\delta_1, \delta_1]$ we obtain

$$\begin{aligned} &\left|I_\tau(\mu_\tau + x; z) - I_\tau(\mu_\tau; z) - (I_{\tau'}(\mu_{\tau'} + x'; z) - I_{\tau'}(\mu_{\tau'}; z))\right| \\ \leq &\left|I_\tau(\mu_\tau + x; z) - I_{\tau'}(\mu_{\tau'} + x'; z)\right| + \left|I_\tau(\mu_\tau; z) - I_{\tau'}(\mu_{\tau'}; z)\right|, \end{aligned}$$

where the second term was discussed above and the first can be handled likewise to conclude

$$\left|I_\tau(\mu_\tau + x; z) - I_{\tau'}(\mu_{\tau'} + x'; z)\right| \leq (|\tau - \tau'| + |x - x'|)\,(C + \delta_1 + |z|) \tag{5.16}$$

with the same C as above. Hence,

$$\begin{aligned} &\left|I_\tau(\mu_\tau + x; z) - I_\tau(\mu_\tau; z) - (I_{\tau'}(\mu_{\tau'} + x'; z) - I_{\tau'}(\mu_{\tau'}; z))\right| \\ &\leq L(z)\,(|\tau - \tau'| + |x - x'|) \end{aligned}$$

is valid with Lipschitz-constant $L(z) = 2\,C + \delta_1 + 2\,|z|$, which is square-integrable by assumption on F. By Example 19.7 in van der Vaart (1998) the bracketing number $N_{[\,]}(\varepsilon, \mathcal{H}_{\delta_1}, \|\cdot\|_{Y,2})$ of $\mathcal{H}_{\delta_1}$ is of order ε^{-2}, so that for the bracketing integral it holds that

$$J_{[\,]}(\varepsilon_n, \mathcal{H}_{\delta_n}, \|\cdot\|_{Y,2}) \leq J_{[\,]}(\varepsilon_n, \mathcal{H}_{\delta_1}, \|\cdot\|_{Y,2}) \to 0$$

as $\varepsilon_n \to 0$. From (5.14), the class $\mathcal{H}_{\delta_n}$ has envelope δ_n, and hence using Corollary 19.35 in van der Vaart (1998), we obtain

$$\mathbb{E}\left[\|\mathbb{G}_n\|_{\mathcal{H}_{\delta_n}}\right] \leq C_1\, J_{[\,]}(\delta_n, \mathcal{H}_{\delta_n}, \|\cdot\|_{Y,2})$$

for some constant $C_1 > 0$. The right hand side converges to zero, such that an application of the Markov inequality ends the proof of (5.3) (this is similar to the reasoning laid out after Definition 2.6). $\square$

For Step 2 in the proof of Theorem 5.1 we have to prove existence of an inverse for ψ_0 and semi-Hadamard differentiability thereof.

Proof of Lemma 5.14. Given $\tau \in [\tau_l, \tau_u]$, by (5.12) and the lower bound in (5.13), the function $x \mapsto I_\tau(x; F)$ is strictly decreasing, and its image is all of $\mathbb{R}$. Hence, for any $u \in \mathbb{R}$ there is a unique x satisfying $I_\tau(x; F) = u$, which shows that ψ_0 is invertible.

Next, for fixed $\varphi \in \ell^\infty([\tau_l, \tau_u])$ the preimage $\big((I_\tau(\cdot; F)\big)^{\mathrm{Inv}}([-\|\varphi\|, \|\varphi\|])$ is by monotonicity an interval $[L_\tau, U_\tau]$, $|L_\tau|, |U_\tau| < \infty$. By (5.11),

$$I_\tau(x; F) = \tau \Big[\int_x^\infty (1 - F(z))\,\mathrm{d}z + \int_{-\infty}^x F(z)\,\mathrm{d}z \Big] - \int_{-\infty}^x F(z)\,\mathrm{d}z,$$

thus the map $\tau \mapsto I_\tau(x; F)$ is increasing, showing $L_{\tau'} \le L_\tau$ and $U_{\tau'} \le U_\tau$ for $\tau \ge \tau'$. This implies that the solution of $u = I_\tau(x; F)$ for $u \in [-\|\varphi\|, \|\varphi\|]$ lies in $[L_{\tau_l}, U_{\tau_u}]$, which means that $\psi_0^{\mathrm{Inv}}(\varphi)$ is bounded. □

Before we turn to the proof of Lemma 5.15, we need the following technical assertions about ψ_0^{Inv}.

5.20 Lemma.

Given $t > 0$ and $\nu \in \ell^\infty([\tau_l, \tau_u])$, we have that

$$\begin{aligned} &t^{-1}(\psi_0^{\mathrm{Inv}}(t\,\nu) - \psi_0^{\mathrm{Inv}}(0))(\tau) \\ &\qquad = \nu(\tau) \left[\tau + (1 - 2\,\tau) \int_0^1 F\big(\mu_\tau + s(\psi_0^{\mathrm{Inv}}(t\,\nu)(\tau) - \mu_\tau)\big)\,\mathrm{d}s \right]^{-1}. \end{aligned} \tag{5.17}$$

In particular, if $\nu_n \in \ell^\infty([\tau_l, \tau_u])$ with $\|\nu_n\| \to 0$, then $\|\psi_0^{\mathrm{Inv}}(\nu_n)(\cdot) - \mu_\cdot\| \to 0$, so that for any $\tau_n, \tau \in [\tau_l, \tau_u]$ with $\tau_n \to \tau$ it holds that

$$\psi_0^{\mathrm{Inv}}(\nu_n)(\tau_n) - \mu_{\tau_n} \to 0. \tag{5.18}$$

The proof of Lemma 5.20 is given in Section 5.9.

We now introduce the following notation for the sequel. Given $\varphi \in \ell^\infty([\tau_l, \tau_u])$ let $c_\varphi \in \ell^\infty([\tau_l, \tau_u])$ be defined by

$$c_\varphi(\tau) = \tau + (1 - 2\,\tau) \int_0^1 F\big(\mu_\tau + s\,\varphi(\tau)\big)\,\mathrm{d}s. \tag{5.19}$$

Proof of Lemma 5.15. Let $t_n \to 0$, $t_n > 0$, and $\varphi_n \in \ell^\infty([\tau_l, \tau_u])$ be a sequence fulfilling $\varphi_n \to \varphi \in \mathcal{C}([\tau_l, \tau_u])$ with respect to d_{hypi} and thus uniformly by Proposition 2.12. From (5.17) and using the notation (5.19) we can write

$$t_n^{-1}(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n) - \psi_0^{\mathrm{Inv}}(0)) = \frac{\varphi_n}{c_{\iota_n}}, \qquad \iota_n(\tau) = \psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \mu_\tau$$

and we need to show that

$$\frac{\varphi_n}{c_{\iota_n}} \to \dot{\psi}_0^{\mathrm{Inv}}(\varphi) = \frac{\varphi}{c_0} \tag{5.20}$$

with respect to d_{hypi}, where $c_0(\tau) = \tau + (1 - 2\,\tau)F(\mu_\tau)$. But, since $\varphi_n \to \varphi$ uniformly and φ is continuous, to obtain (5.20) if suffices by Lemma 2.13, i) and ii), to show that $c_{\kappa_n} \to c_0$ under d_{hypi}. To this end, by Lemma A.4, Bücher et al. (2014) and Lemma 2.13, i), it suffices to show that the convergence

$$h_n(\tau) = \int_0^1 F\big(\mu_\tau + s(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \mu_\tau)\big)\,\mathrm{d}s \to h(\tau) := F(\mu_\tau) \tag{5.21}$$

with respect to d_{hypi} is true, for which we shall use Corollary A.7 in Bücher et al. (2014). Let

$$\mathbb{T} = [\tau_l, \tau_u], \qquad \mathcal{S} = \mathbb{T} \setminus \{\tau \in \tau \mid F \text{ is not continuous in } \mu_\tau\}, \tag{5.22}$$

so that $\mathcal{S}$ is dense in $\mathbb{T}$ and $h|_{\mathcal{S}}$ is continuous. Observe that $\mu_{\tau_l}, \mu_{\tau_u} \in \mathcal{S}$ holds by assumption. Further, note that due to the monotonicity of $\tau \mapsto \mu_\tau$ (Holzmann and Klar, 2016, Proposition 1) as well as monotonicity and right-continuity of F it holds that

$$\lim_{\tau' \nearrow \tau} F(\mu_{\tau'}) = F(\mu_\tau -) \text{ and } \lim_{\tau' \searrow \tau} F(\mu_{\tau'}) = F(\mu_\tau).$$

Thus, using the notation from Bücher et al. (2014), Appendix A.2, we have that

$$(h|_{\mathcal{S}})_\wedge^{\mathcal{S}:\mathbb{T}} = h_\wedge = F(\mu_\cdot -) \quad \text{and} \quad (h|_{\mathcal{S}})_\vee^{\mathcal{S}:\mathbb{T}} = h_\vee = h,$$

where the first equalities additionally need the discussion in Bücher et al. (2014), Appendix A.2, and the second equalities Lemma 2.9, ii). If we show that

i) for all $\tau_n, \tau \in [\tau_l, \tau_u]$ with $\tau_n \to \tau$ it holds that $\liminf_n h_n(\tau_n) \geq F(\mu_\tau -)$ and

ii) for all $\tau_n, \tau \in [\tau_l, \tau_u]$ with $\tau_n \to \tau$ it holds that $\limsup_n h_n(\tau_n) \leq F(\mu_\tau)$,

Corollary A.7 in Bücher et al. (2014) implies (5.21), which concludes the proof of the convergence in (5.20). Therefore choose any sequence $\tau_n \to \tau$.

Concerning i), we compute that

$$\begin{aligned} F(\mu_\tau -) &\leq \int_0^1 \liminf_n F\big(\mu_{\tau_n} + s\,(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau_n) - \mu_{\tau_n})\big)\,\mathrm{d}s \\ &\leq \liminf_n \int_0^1 F\big(\mu_{\tau_n} + s\,(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau_n) - \mu_{\tau_n})\big)\,\mathrm{d}s = \liminf_n h_n(\tau_n), \end{aligned}$$

where the first inequality follows from (5.18) and the fact that $F(\mu_\tau -) \leq F(\mu_\tau)$, and the second inequality follows from Fatou's lemma. For ii) we argue analogously as

$$\begin{aligned} F(\mu_\tau) &\geq \int_0^1 \limsup_n F\big(\mu_{\tau_n} + s\,(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau_n) - \mu_{\tau_n})\big)\,\mathrm{d}s \\ &\geq \limsup_n \int_0^1 F\big(\mu_{\tau_n} + s\,(\psi_0^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau_n) - \mu_{\tau_n})\big)\,\mathrm{d}s = \limsup_n h_n(\tau_n). \end{aligned}$$

This concludes the proof of the lemma. □

5.6.2.2 Details for the proof of Theorem 5.4

Here we turn to the proof of the (strong) consistency of the bootstrap procedure for the expectile process. We let $\psi_n^*(\varphi)(\tau) = -I_\tau(\varphi(\tau); F_n^*)$, $\varphi \in \ell^\infty([\tau_l, \tau_u])$, and denote by P_n^* the conditional law of $Y_1^*, \ldots, Y_n^*$ given $Y_1, \ldots, Y_n$, and by $\mathbb{E}_n^*$ the expectation under P_n^*.

The next lemma includes the assertions paralleling the first step in the proof of Theorem 5.1; the proof is contained in Section 5.9.

5.21 Lemma.

We have, almost surely, conditionally on $Y_1, Y_2, \ldots$, the following statements.

i) If $Y \in \mathcal{L}_1$, then

$$\|\mu_{\tau,n}^* - \widehat{\mu}_{\tau,n}\| = o_{\mathrm{P}_n^*}(1). \tag{5.23}$$

Now assume $Y \in \mathcal{L}_2$.

ii) Weakly in $(\ell^\infty([\tau_l, \tau_u]), \|\cdot\|)$ it is true that

$$\sqrt{n}(\psi_n^*(\widehat{\mu}_{\cdot,n}) - \psi_n(\widehat{\mu}_{\cdot,n})) \rightsquigarrow Z. \tag{5.24}$$

with Z as in Theorem 5.1.

iii) For every sequence $\delta_n \to 0$ it holds that

$$\sup_{\|\varphi\| \le \delta_n} \sqrt{n} \,\big\|\psi_n^*(\widehat{\mu}_{\cdot,n} + \varphi)(\cdot) - \psi_n(\widehat{\mu}_{\cdot,n} + \varphi)(\cdot) - \big[\psi_n^*(\widehat{\mu}_{\cdot,n})(\cdot) - \psi_n(\widehat{\mu}_{\cdot,n})(\cdot)\big]\big\| = o_{\mathrm{P}_n^*}(1). \tag{5.25}$$

iv) Weakly in $(\ell^\infty([\tau_l, \tau_u]), \|\cdot\|)$ we have that

$$\sqrt{n}(\psi_n(\mu_{\cdot,n}^*) - \psi_n(\widehat{\mu}_{\cdot,n})) \rightsquigarrow Z.. \tag{5.26}$$

Additionally, we need a kind of asymptotic semi-Hadamard differentiability.

5.22 Lemma.

The map ψ_n is invertible, and if $t_n \to 0$, $t_n > 0$, $\varphi_n \in \ell^\infty([\tau_l, \tau_u])$, $\varphi \in \mathcal{C}[\tau_l, \tau_u]$, where $\varphi_n \to \varphi$ with respect to d_{hypi}, we almost surely, conditionally on $Y_1, Y_2, \ldots$, have the hypi-convergence

$$t_n^{-1}(\psi_n^{\mathrm{Inv}}(t_n\varphi_n) - \psi_n^{\mathrm{Inv}}(0)) \to \dot{\psi}_0^{\mathrm{Inv}}(\varphi). \tag{5.27}$$

Proof. The first part follows from Lemma 5.14 with F in ψ_0 replaced by F_n in ψ_n as no specific assumptions on F were used in that lemma.

For (5.27), with the same calculations as for Lemma 5.20 we obtain the representation

$$\begin{aligned} &t_n^{-1}\big(\psi_n^{\mathrm{Inv}}(t_n\varphi_n) - \psi_n^{\mathrm{Inv}}(0)\big)(\tau) \\ &\quad = \varphi_n(\tau)\left[\tau + (1-2\,\tau)\int_0^1 F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s\right]^{-1}, \end{aligned}$$

and we have to prove hypi-convergence thereof to $\dot{\psi}_0^{\mathrm{Inv}}(\varphi)$.

By the same reductions as in the proof of Lemma 5.15, it suffices to prove the hypi-convergence of

$$h_n(\tau) = \int_0^1 F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s$$

to $h(\tau) = F(\mu_\tau)$ for almost every sequence $Y_1, Y_2, \ldots$. To this end, observe that for any $s \in [0,1]$ the sequence $\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)$ converges to μ_τ uniformly over τ almost surely by the same arguments as in Lemma 5.20. Since μ_τ is continuous in τ and $\widehat{\mu}_{\tau_n,n}$ is uniformly strongly consistent, for any sequence $\tau_n \to \tau$ the almost sure convergence $\widehat{\mu}_{\tau_n,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau_n) - \widehat{\mu}_{\tau_n,n}\big) \to \mu_\tau$ follows. By adding and subtracting the term $F_n(\widehat{\mu}_{\tau,n} + s(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}))$ and using Lemma 2.24, we now can estimate

$$\begin{aligned} F(\mu_\tau-) &\le \int_0^1 \liminf_n \Bigg(F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\Bigg) \\ &\qquad + \limsup_n \Bigg(F\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big) \\ &\qquad\quad - F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\Bigg)\mathrm{d}s \\ &\le \int_0^1 \liminf_n F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s + \limsup_n \|F_n - F\|_{\mathbb{R}} \\ &\le \liminf_n \int_0^1 F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s = \liminf_n h_n(\tau_n) \end{aligned}$$

almost surely, where the "$\limsup$"-part vanishes due to the Glivenko-Cantelli-Theorem for the empirical distribution function. Similarly, we almost surely have

$$\begin{aligned} F(\mu_\tau) &\ge \int_0^1 \limsup_n F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s + \liminf_n \|F_n - F\|_{\mathbb{R}} \\ &\ge \limsup_n \int_0^1 F_n\big(\widehat{\mu}_{\tau,n} + s\big(\psi_n^{\mathrm{Inv}}(t_n\,\varphi_n)(\tau) - \widehat{\mu}_{\tau,n}\big)\big)\,\mathrm{d}s = \limsup_n h_n(\tau_n). \end{aligned}$$

As in the proof of Lemma 5.15 we conclude with Corollary A.7, Bücher et al. (2014). □

With the assertions so far we can conclude Theorem 5.4.

Proof of Theorem 5.4. Set $t_n = n^{-1/2}$ and define the function

$$g_n(\varphi) = t_n^{-1}\,(\psi_n^{\mathrm{Inv}}(t_n\,\varphi) - \psi_n^{\mathrm{Inv}}(0)), \quad \varphi \in \ell^\infty([\tau_l, \tau_u]).$$

Then, from (5.27) the hypi-convergence $g_n(\varphi_n) \to \dot{\psi}_0^{\mathrm{Inv}}(\varphi)$ holds almost surely, whenever $\varphi_n \in \ell^\infty([\tau_l, \tau_u])$, $\varphi \in \mathcal{C}[\tau_l, \tau_u]$ such that $\varphi_n \to \varphi$ with respect to d_{hypi}. In addition, $\sqrt{n}(\psi_n(\mu^*_{\cdot,n}) - \psi_n(\widehat{\mu}_{\cdot,n})) \rightsquigarrow Z_\cdot$ with respect to the supremum-norm, conditionally on $Y_1, Y_2, \ldots$ almost surely, by (5.24), where Z is continuous almost surely. Hence, the convergence is also valid with respect to d_{hypi}, such that

$$\sqrt{n}(\mu^*_{\cdot,n} - \widehat{\mu}_{\cdot,n}) = g_n\big(\sqrt{n}(\psi_n(\mu^*_{\cdot,n}) - \psi_n(\widehat{\mu}_{\cdot,n}))\big) \rightsquigarrow \dot{\psi}_0^{\mathrm{Inv}}(Z_\cdot)(\cdot)$$

holds weakly in $\big(\ell^\infty([\tau_l, \tau_u]), \mathrm{d}_{hypi}\big)$, conditionally on $Y_1, Y_2, \ldots$ almost surely, by using the extended continuous mapping theorem, Theorem B.3, in Bücher et al. (2014). □

5.6.2.3 Details for the proof of Theorem 5.6

Here we give the remaining part of the proof of Theorem 5.6. More precisely, we determine the shape of the left- and right-sided derivatives of F^{Inv} and show uniform semi-Hadamard differentiability of the map $\Psi_0^{\mathrm{Inv}}(\nu)(\alpha) = F^{\mathrm{Inv}}(\alpha + \nu(\alpha))$.

Proof of Lemma 5.5. Choose $\alpha \in (F(q_{\alpha_l}) - \delta, F(q_{\alpha_u}) + \delta)$ and assume first that F is continuous in q_α. In this case we can rewrite the left-sided difference quotient as

$$\frac{F^{\mathrm{Inv}}(\alpha) - F^{\mathrm{Inv}}(\alpha - h)}{h} = \left(\frac{F\big(F^{\mathrm{Inv}}(\alpha)\big) - F\big(F^{\mathrm{Inv}}(\alpha - h)\big)}{F^{\mathrm{Inv}}(\alpha) - F^{\mathrm{Inv}}(\alpha - h)}\right)^{-1},$$

where $h > 0$. As F^{Inv} is continuous and monotonically increasing, $F^{\mathrm{Inv}}(\alpha - h) \nearrow F^{\mathrm{Inv}}(\alpha)$ is valid, such that the above right hand side converges to $\big(\partial^-(F)(F^{\mathrm{Inv}}(\alpha))\big)^{-1}$ if $\partial^-(F)(F^{\mathrm{Inv}}(\alpha))$ is bounded. If $\partial^-(F)(F^{\mathrm{Inv}}(\alpha))$ is unbounded, the fraction above converges to 0, which is how $\big(\partial^-(F)(F^{\mathrm{Inv}}(\alpha))\big)^{-1}$ must be read in this case. Analogue arguments yield the convergence of the right-sided difference quotient of F^{Inv} in α to $\big(\partial^+(F)(F^{\mathrm{Inv}}(\alpha))\big)^{-1}$. This proves i).

Next, we consider one of the y_i where F jumps and choose α_i with $y_i = F^{\mathrm{Inv}}(\alpha_i)$; for simplicity we suppress the index and write $\alpha = \alpha_i$ and $y_i = q_\alpha$. Let us choose a sequence $h \searrow 0$ and consider the left-sided difference quotient

$$\frac{F^{\mathrm{Inv}}(\alpha) - F^{\mathrm{Inv}}(\alpha - h)}{h}.$$

We choose h small enough, such that $F^{\mathrm{Inv}}(\alpha - h) \neq y_i$ for all $i \in \{1, \ldots, r\}$. This is possible, since there are only r values, where F is discontinuous and thus they are isolated from each other (there exist pairwise disjoint open neighbourhoods around the y_i). Then we know that F is continuous in $F^{\text{Inv}}(\alpha - h)$ for every small h.

Now, we first suppose that $\alpha \in \big(F(q_\alpha -), F(q_\alpha)\big)$, so $(\alpha - \Delta, \alpha + \Delta) \subset \big(F(q_\alpha -), F(q_\alpha)\big)$ for some small $\Delta > 0$. For every $\alpha' \in \big(F(q_\alpha -), F(q_\alpha)\big)$ it holds that $F^{\text{Inv}}(\alpha') = F^{\text{Inv}}(\alpha)$, especially this is valid for every $\alpha' \in (\alpha - \Delta, \alpha + \Delta)$. But, for h small enough, we know that $\alpha - h$, $\alpha + h \in (\alpha - \Delta, \alpha + \Delta)$, yielding the equality

$$\frac{F^{\text{Inv}}(\alpha) - F^{\text{Inv}}(\alpha - h)}{h} = 0 \quad \text{and} \quad \frac{F^{\text{Inv}}(\alpha + h) - F^{\text{Inv}}(\alpha)}{h} = 0$$

This shows $\partial^{\pm}\big(F^{\text{Inv}}\big)(\alpha) = 0$; note that in the present case $\partial^{-}(F)(q_\alpha) = \infty$ is true, hence the representation $\partial^{\pm}\big(F^{\text{Inv}}\big)(\alpha) = \big(\partial^{-}(F)(q_\alpha)\big)^{-1}$ can be used again. This shows ii).

For iii) assume that $\alpha = F(q_\alpha)$. Then we choose $\Delta > 0$ fulfilling $(\alpha - \Delta, \alpha) \subset \big(F(q_\alpha -), F(q_\alpha)\big)$, which leads to $\partial^{-}\big(F^{\text{Inv}}\big)(\alpha) = \big(\partial^{-}(F)(q_\alpha)\big)^{-1}$ with the same reasoning as before. On the other hand, the right-sided difference quotient can be translated to

$$\begin{aligned}
\frac{F^{\text{Inv}}(\alpha + h) - F^{\text{Inv}}(\alpha)}{h} &= \left(\frac{\alpha + h - \alpha}{F^{\text{Inv}}(\alpha + h) - F^{\text{Inv}}(\alpha)}\right)^{-1} \\
&= \left(\frac{F\big(F^{\text{Inv}}(\alpha + h)\big) - F\big(F^{\text{Inv}}(\alpha)\big)}{F^{\text{Inv}}(\alpha + h) - F^{\text{Inv}}(\alpha)}\right)^{-1},
\end{aligned}$$

where the last fraction converges to $\big(\partial^{+}(F)(q_\alpha)\big)^{-1}$ as asserted.

In order to show iv), assume that $\alpha = F(q_\alpha -)$. In this case we can choose a $\Delta > 0$ with $(\alpha, \alpha + \Delta) \subset \big(F(q_\alpha -), F(q_\alpha)\big)$, to obtain $\partial^{+}\big(F^{\text{Inv}}\big)(\alpha) = \big(\partial^{-}(F)(q_\alpha)\big)^{-1}$ again. For the left-sided derivative we know by the (extended) Theorem of Rolle, that there is a $\xi_h \in \big(\alpha - h, \alpha\big)$ with

$$\partial^{-}(F^{\text{Inv}})(\xi_h) \wedge \partial^{+}(F^{\text{Inv}})(\xi_h) \leq \frac{F^{\text{Inv}}(\alpha) - F^{\text{Inv}}(\alpha - h)}{h} \leq \partial^{-}(F^{\text{Inv}})(\xi_h) \vee \partial^{+}(F^{\text{Inv}})(\xi_h). \tag{5.28}$$

For h small enough, we know that F does not jump in $F^{\text{Inv}}(\xi_h)$ and hence we can use the results shown so far to deduce

$$\begin{aligned}
&\big(\partial^{-}(F)(F^{\text{Inv}}(\xi_h))\big)^{-1} \wedge \big(\partial^{+}(F)(F^{\text{Inv}}(\xi_h))\big)^{-1} \\
&\qquad \leq \frac{F^{\text{Inv}}(\alpha) - F^{\text{Inv}}(\alpha - h)}{h} \\
&\qquad\quad \leq \big(\partial^{-}(F)(F^{\text{Inv}}(\xi_h))\big)^{-1} \vee \big(\partial^{+}(F)(F^{\text{Inv}}(\xi_h))\big)^{-1}.
\end{aligned}$$

We observe that $F^{\text{Inv}}(\xi_h) \nearrow F^{\text{Inv}}(\alpha)$ as $h \searrow 0$, and as both one-sided derivatives of F have existing limits in every point by Assumption [C], the lower and upper bounds above converge to the minimum and maximum of the values $\big(\partial^{-}(F)(F^{\text{Inv}}(\alpha)-)\big)^{-1}$ and

$\big(\partial^+(F)(F^{\text{Inv}}(\alpha)-)\big)^{-1}$, respectively. Since the points, where F is differentiable, form a dense set, we can choose a sequence $y_s \nearrow F^{\text{Inv}}(\alpha)$ such that $F'(y_s)$ exists. For this sequence it holds that

$$\partial^-(F)(F^{\text{Inv}}(\alpha)-) = \lim_{s\to\infty} \partial^-(F)(y_s) = \lim_{s\to\infty} \partial^+(F)(y_s) = \partial^+(F)(F^{\text{Inv}}(\alpha)-)$$

where all values are in the interval $[c, \infty]$. Thus, the upper and lower bound in (5.28) both converge to $\big(\partial^{\pm}(F)(F^{\text{Inv}}(\alpha)-)\big)^{-1} \in [0, \frac{1}{c}]$, showing the asserted convergence of the left-sided difference quotient of F^{Inv} in α. ☐

Using the existence of the one-sided derivatives of F^{Inv}, we can show the semi-Hadamard differentiability remaining to conclude Theorem 5.6.

Proof of Lemma 5.16. Let $t_n \searrow 0$, $\alpha_n \to \alpha \in [\alpha_l, \alpha_u]$ and $\varphi_n, \nu_n \in \ell^\infty[\alpha_l, \alpha_u]$ with $\varphi_n \to \varphi \in \mathcal{C}[\alpha_l, \alpha_u]$ and $\nu_n \to 0$ with respect to d_{hypi}. Then $\varphi_n(\alpha_n) \to \varphi(\alpha)$ and $\nu_n(\alpha_n) \to 0$ holds by Proposition 2.1, Bücher et al. (2014). We have to deal with the limes inferior and superior of

$$t_n^{-1}\Big[F^{\text{Inv}}(\alpha_n + \nu_n(\alpha_n) + t_n\,\varphi_n(\alpha_n)) - F^{\text{Inv}}(\alpha_n + \nu_n(\alpha_n))\Big]$$

which can be rewritten as

$$\varphi_n(\alpha_n)\,\frac{F^{\text{Inv}}(\alpha_n + \nu_n(\alpha_n) + t_n\,\varphi_n(\alpha_n)) - F^{\text{Inv}}(\alpha_n + \nu_n(\alpha_n))}{t_n\,\varphi_n(\alpha_n)}.$$

By Lemma 2.13 we only have to deal with the accumulation points of the fraction above, which we call $H_n(\alpha_n)$ for convenience. Similar as in the case for the expectile process, we utilize Corollary A.7, Bücher et al. (2014), for which we define

$$\mathbb{T} = [\alpha_l, \alpha_u] \quad \text{and} \quad \mathcal{S} = \mathbb{T} \setminus \big\{\alpha \in \mathbb{T} \mid \partial^-(F^{\text{Inv}}) \text{ is not continuous in } \alpha\big\}. \tag{5.29}$$

Since $\alpha \mapsto \partial^-(F^{\text{Inv}})(\alpha)$ is càdlàg or làdcàg in every point by the former lemma, we know that $\mathcal{S}$ is dense in $\mathbb{T}$ with $\alpha_l, \alpha_u \in \mathcal{S}$ by Assumption [C] and, by definition of $\mathcal{S}$, $\partial^-(F^{\text{Inv}})|_{\mathcal{S}}$ is continuous. Again due to the properties of $\alpha \mapsto \partial^-(F^{\text{Inv}})(\alpha)$ it further holds that

$$\begin{aligned}
\big(\partial^-(F^{\text{Inv}})|_{\mathcal{S}}\big)_{\wedge}^{\mathcal{S}:\mathbb{T}} &= \min\{\partial^-(F^{\text{Inv}})(\alpha-),\, \partial^-(F^{\text{Inv}})(\alpha+)\} = \big(\partial^-(F^{\text{Inv}})\big)_{\wedge}(\alpha) \quad \text{and} \\
\big(\partial^-(F^{\text{Inv}})|_{\mathcal{S}}\big)_{\vee}^{\mathcal{S}:\mathbb{T}} &= \max\{\partial^-(F^{\text{Inv}})(\alpha-),\, \partial^-(F^{\text{Inv}})(\alpha+)\} = \big(\partial^-(F^{\text{Inv}})\big)_{\vee}(\alpha),
\end{aligned}$$

where we again used the notation and discussion in Bücher et al. (2014), Appendix A.2, and Lemma 2.9. So we only need to show

$$\big(\partial^-(F^{\text{Inv}})\big)_{\wedge}(\alpha) \le \liminf_n H_n(\alpha_n) \quad \text{and} \quad \big(\partial^-(F^{\text{Inv}})\big)_{\vee}(\alpha) \ge \limsup_n H_n(\alpha_n).$$

For this we want to use the extended Theorem of Rolle, which allows us to bound the difference quotient in a suitable way. Therefore, observe that

$$\alpha_n + \nu_n(\alpha_n) + t_n\,\varphi_n(\alpha_n),\ \alpha_n + \nu(\alpha_n) \in [\alpha - \|\nu_n\| - t_n\,\|\varphi_n\|,\ \alpha + \|\nu_n\| + t_n\,\|\varphi_n\|],$$

where the length of the above interval converges to zero for n to infinity. Rolle's Theorem yields a $\xi_n \in [\alpha - \|\nu_n\| - t_n\,\|\varphi_n\|,\ \alpha + \|\nu_n\| + t_n\,\|\varphi_n\|]$ with

$$\partial^-(F^{\mathrm{Inv}})(\xi_n) \wedge \partial^+(F^{\mathrm{Inv}})(\xi_n) \le H_n(\alpha_n) \le \partial^-(F^{\mathrm{Inv}})(\xi_n) \vee \partial^+(F^{\mathrm{Inv}})(\xi_n);$$

note that $\xi_n \to \alpha$ for n to infinity. By the properties of the lower- and upper-semicontinuous hulls, this is equivalent to

$$\big(\partial^-(F^{\mathrm{Inv}})\big)_\wedge(\xi_n) \le H_n(\alpha_n) \le \big(\partial^+(F^{\mathrm{Inv}})\big)_\vee(\xi_n)$$

and again using the properties of the hulls, we conclude

$$\big(\partial^-(F^{\mathrm{Inv}})\big)_\wedge(\alpha) \le \liminf_n H_n(\alpha_n) \quad \text{and} \quad \big(\partial^-(F^{\mathrm{Inv}})\big)_\vee(\alpha) \ge \limsup_n H_n(\alpha_n).$$

Now, Corollary A.7, Bücher et al. (2014), yields the assertion. □

5.7 Proof of Theorem 5.11

In this section we sketch how to prove the semi-Hadamard differentiability of $h \mapsto h^{\mathrm{Inv}}$ as formulated in Theorem 5.11. The stated semi-derivative was $-\partial^-(F^{\mathrm{Inv}})\,Z \circ F^{\mathrm{Inv}}$, where $Z \in \mathbb{W} = \{\varphi \in \mathcal{D}([q_{\alpha_l}, q_{\alpha_u}]) \,|\, \varphi \text{ jumps at most if } F \text{ jumps}\}$. Using the next statement and $(-h)_\wedge = -h_\vee$, we can determine the semicontinuous hulls of this limit process. The assertion is a combination of Corollary 2.10 and Lemma 2.13.

5.23 Corollary.

Let Assumption [C] hold for the distribution function F and $\varphi \in \mathbb{W}$. For $\alpha \in [\alpha_l, \alpha_u]$ the following assertions are true.

i) If $\alpha \notin [F(q_\alpha -), F(q_\alpha))$, then it holds that

$$\Big(\varphi \circ F^{\mathrm{Inv}}\,\partial^-(F^{\mathrm{Inv}})\Big)_\wedge(\alpha) = \varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha)\,\mathbb{1}\,(\varphi(q_\alpha) > 0) + \varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\,\mathbb{1}\,(\varphi(q_\alpha) < 0)$$

and

$$\Big(\varphi \circ F^{\mathrm{Inv}}\,\partial^-(F^{\mathrm{Inv}})\Big)_\vee(\alpha) = \varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\,\mathbb{1}\,(\varphi(q_\alpha) > 0) + \varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha)\,\mathbb{1}\,(\varphi(q_\alpha) < 0)$$

with

$$\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha) = \min\left\{\left(\partial^-(F)(q_\alpha)\right)^{-1},\ \left(\partial^+(F)(q_\alpha)\right)^{-1}\right\} \quad \textit{and}$$
$$\partial^-(F^{\mathrm{Inv}})_\vee(\alpha) = \max\left\{\left(\partial^-(F)(q_\alpha)\right)^{-1},\ \left(\partial^+(F)(q_\alpha)\right)^{-1}\right\}.$$

ii) Assume $\alpha \in (F(q_\alpha-), F(q_\alpha))$. The map $\partial^-(F^{\mathrm{Inv}})$ is constant on the latter interval, more precisely it is valid that $\partial^-(F^{\mathrm{Inv}})|_{(F(q_\alpha-),F(q_\alpha))} = 0$. Especially, we have that

$$\left(\varphi \circ F^{\mathrm{Inv}}\, \partial^-(F^{\mathrm{Inv}})\right)_\wedge(\alpha) = 0 = \left(\varphi \circ F^{\mathrm{Inv}}\, \partial^-(F^{\mathrm{Inv}})\right)_\vee(\alpha).$$

iii) Last, let $\alpha = F(q_\alpha-)$. Here, it holds that

$$\left(\varphi \circ F^{\mathrm{Inv}}\, \partial^-(F^{\mathrm{Inv}})\right)_\wedge(\alpha) = \varphi(q_\alpha-)\, \partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\, \mathbb{1}\left(\varphi(q_\alpha-) < 0\right)$$

and

$$\left(\varphi \circ F^{\mathrm{Inv}}\, \partial^-(F^{\mathrm{Inv}})\right)_\vee(\alpha) = \varphi(q_\alpha-)\, \partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\, \mathbb{1}\left(\varphi(q_\alpha-) > 0\right),$$

where $\partial^-(F^{\mathrm{Inv}})_\vee(\alpha) = \left(\partial^-(F)(q_\alpha-)\right)^{-1}(\alpha)$.

Proof. Note that $\varphi \circ F^{\mathrm{Inv}}$ is continuous for $\alpha \neq F(q_\alpha-)$. Thus, the assertions in i) and ii) follow from Lemma 2.13, i), and Lemma 5.5, i) and ii). Assertion iii) is implied by Corollary 2.10, ii), and Lemma 5.5, iii). □

The last preparation for proving Theorem 5.11 is the following.

5.24 Lemma.

Let a_n be a sequence and a_{n_j}, a_{n_k} a partition of the original sequence (one is allowed be finite or empty). Then it holds that

$$\limsup_n a_n = \left(\limsup_j a_{n_j}\right) \vee \left(\limsup_k a_{n_k}\right) \quad \textit{and}$$
$$\liminf_n a_n = \left(\liminf_j a_{n_j}\right) \wedge \left(\liminf_k a_{n_k}\right),$$

where we use $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$.

Proof. Consider the first equality. It certainly holds that the left hand side is greater than or equal to the right hand side, as $\{a_{n_j}\}$, $\{a_{n_k}\} \subset \{a_n\}$ is valid. On the other hand, by the properties of the limes superior there is a subsequence a_{n_l} for which it holds that

$$\limsup_n a_n = \lim_l a_{n_l}.$$

An infinite number of the members of this sequence must either be in the fixed sequence a_{n_k} or a_{n_j}, which shows $\limsup_n a_n \leq \Big(\limsup_j a_{n_j}\Big) \vee \Big(\limsup_k a_{n_k}\Big)$. Thus, the stated equality must be true.

The assertion about the limes inferior follows similarly. □

Proof of Theorem 5.11. Let $0 < t_n \to 0$ and $\varphi_n \in \mathcal{D}([q_{\alpha_l}, q_{\alpha_u}])$ fulfilling $\varphi_n \longrightarrow \varphi \in \mathbb{W}$ with respect to $\|\cdot\|$ as well as $F + t_n\,\varphi_n \in \mathcal{D}_0$. We need to show

$$\mathrm{d}_{hypi}\Big(t_n^{-1}(\Phi(F + t_n\,\varphi_n) - \Phi(F)),\, -\varphi \circ F^{\mathrm{Inv}}\,\partial^-(F^{\mathrm{Inv}})\Big) \longrightarrow 0,$$

that is we have to consider the accumulation points of

$$t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \tag{5.30}$$

for any sequence $\alpha_n \to \alpha$ with α_n, $\alpha \in [\alpha_l, \alpha_u]$. The asserted derivative $\dot{\Phi}(\varphi)$ is càdlàg or làdcàg in every point and the lower- and upper-semicontinuous hulls can be represented by an appropriate product of $\varphi(F^{\mathrm{Inv}}(\cdot)\pm)$ and $\partial^-(F^{\mathrm{Inv}})(\cdot\pm)$, see Corollary 5.23. Especially, we can again use Corollary A.7, Bücher et al. (2014), and only need to consider the limes inferior and superior of the expression in (5.30) in that we have to show

$$\liminf_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \geq \big(\dot{\Phi}(\varphi)\big)_\wedge(\alpha) \quad \text{and}$$
$$\limsup_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \leq \big(\dot{\Phi}(\varphi)\big)_\vee(\alpha).$$

After some introductory remarks, we distinguish four cases:

Case 1) $\alpha \in \big(F(y_i-), F(y_i)\big)$ for some $i \in \{1, \ldots, r\}$;

Case 2) $\alpha \notin [F(y_i-), F(y_i)]$ for every $i \in \{1, \ldots, r\}$;

Case 3) $\alpha = F(y_i-)$ for some $i \in \{1, \ldots, r\}$;

Case 4) $\alpha = F(y_i)$ for some $i \in \{1, \ldots, r\}$.

An important observation is that $\|t_n \varphi_n\| = \Delta_n \to 0$ and $F + t_n \varphi_n \in [\![F - \Delta_n,\, F + \Delta_n]\!]$, such that

$$\begin{aligned}
F^{\mathrm{Inv}}(\alpha_n - \Delta_n) = \big(F + \Delta_n\big)^{\mathrm{Inv}}(\alpha_n) \\
\leq \big(F + t_n \varphi_n\big)^{\mathrm{Inv}}(\alpha_n) \\
\leq \big(F - \Delta_n\big)^{\mathrm{Inv}}(\alpha_n) = F^{\mathrm{Inv}}(\alpha_n + \Delta_n)
\end{aligned} \tag{5.31}$$

is valid. Additionally, since $\alpha_n + \Delta_n \to \alpha$ it holds that $F^{\mathrm{Inv}}(\alpha_n \pm \Delta_n) \in (q_{\alpha_l} - \varepsilon, q_{\alpha_u} + \varepsilon)$ for n big enough, where ε is given in Assumption [C]. Without loss of generality we assume that this inclusion holds for all n.

This immediately enables us to handle case 1).

Case 1) Let $\alpha \in \big(F(q_\alpha-), F(q_\alpha)\big)$. The assertion is that

$$t_n^{-1}\Big(\big(F + t_n \varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) = 0 \tag{5.32}$$

for n large enough, such that the limes inferior and superior of (5.30) coincide. First, observe that since $\alpha_n \to \alpha$, it holds that $\alpha_n \in \big(F(q_\alpha-), F(q_\alpha)\big)$ for n big enough, such that $F^{\mathrm{Inv}}(\alpha_n) = q_\alpha$ eventually. Using Δ_n above, both sequences $\alpha_n \pm \Delta_n$ converge to α, such that $\alpha_n \pm \Delta_n \in \big(F(q_\alpha-), F(q_\alpha)\big)$ again by choosing n big enough. Then the lower and upper bound in (5.31) both reduce to $F^{\mathrm{Inv}}(\alpha) = q_\alpha$, such that $\big(F + t_n \varphi_n\big)^{\mathrm{Inv}}(\alpha_n) = q_\alpha$ as well. This shows (5.32), and using Corollary 5.23, ii), finishes case 1).

For the remaining cases we start with a general observation. Due to the definition of quantiles (Definition 1.2) for any $\varepsilon_n > 0$ it holds that

$$\big(F + t_n \varphi_n\big)\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) \leq \alpha_n \leq \big(F + t_n \varphi_n\big)\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n)\big).$$

This can be reorganized to

$$F\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) - \alpha_n \leq -t_n \varphi_n\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) \tag{5.33}$$

on the one hand and

$$-\, t_n \varphi_n\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) \leq F\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) - \alpha_n \tag{5.34}$$

on the other hand. As $\alpha_n \leq F\big(F^{\mathrm{Inv}}(\alpha_n)\big)$, (5.33) yields

$$F\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) - F\big(F^{\mathrm{Inv}}(\alpha_n)\big) \leq -t_n \varphi_n\big((F + t_n \varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big).$$

Expanding the left hand side with $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n - F^{\mathrm{Inv}}(\alpha_n)$ and reorganizing finally gives

$$t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \\ \le -\varphi_n\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big)\left(\frac{F\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) - F\big(F^{\mathrm{Inv}}(\alpha_n)\big)}{\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n - F^{\mathrm{Inv}}(\alpha_n)}\right)^{-1} + \frac{\varepsilon_n}{t_n}. \tag{5.35}$$

Note that the "big" fraction is positive due to the monotonicity of F. Further, observe that ε_n was arbitrary so far, but for this inequality to contain information we must choose ε_n such that $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n - F^{\mathrm{Inv}}(\alpha_n) \neq 0$.

In order to reshape (5.34) in a similar way, notice that $\alpha_n \ge F\big(F^{\mathrm{Inv}}(\alpha_n) - \delta_n\big)$ for any $\delta_n > 0$. As a first step, with (5.34) we thus deduce

$$-t_n\,\varphi_n\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) \le F\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) - F\big(F^{\mathrm{Inv}}(\alpha_n) - \delta_n\big).$$

We expand the right hand side with $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n) + \delta_n$, where we choose δ_n such that the latter term is non-zero, and reorganize the resulting inequality to end with

$$t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \\ \ge -\varphi_n\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big)\left(\frac{F\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) - F\big(F^{\mathrm{Inv}}(\alpha_n) - \delta_n\big)}{\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n) + \delta_n}\right)^{-1} - \frac{\delta_n}{t_n}. \tag{5.36}$$

Again the "big" fraction is positive because of the monotonicity of F.

Next, we observe that due to the continuity of the map $\alpha \mapsto F^{\mathrm{Inv}}(\alpha)$ and (5.31) the convergences $(F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) \to q_\alpha$ and $F^{\mathrm{Inv}}(\alpha_n) \to q_\alpha$ are valid. In the following we must ensure $\varepsilon_n,\, \delta_n = o(t_n)$, in addition we want to choose ε_n and δ_n small enough such that

$$\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n,\; F^{\mathrm{Inv}}(\alpha_n) - \delta_n \in (q_{\alpha_l} - \varepsilon, q_{\alpha_u} + \varepsilon)$$

holds, where ε is given in Assumption [C]. This is always possible because of the convergences of $(F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)$ and $F^{\mathrm{Inv}}(\alpha_n)$: At least for n big enough we can choose $\varepsilon_n,\, \delta_n \le t_n^2$; if necessary, we diminish ε_n and δ_n again. Thus, we can forget about $t_n^{-1}\,\varepsilon_n$ in (5.35) and $t_n^{-1}\,\delta_n$ in (5.36) in the following.

Case 2) In this case it holds that

$$\varphi_n\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big),\, \varphi_n\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big) \longrightarrow \varphi(q_\alpha)$$

by the uniform convergence of φ_n and continuity of φ in q_α.

Further, in both (5.35) and (5.36) the remaining fractions (the "big" ones) are the reciprocal of a difference quotient in F over the intervals

$$\big[\big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\wedge F^{\mathrm{Inv}}(\alpha_n),\ \big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\vee F^{\mathrm{Inv}}(\alpha_n)\big]$$

and

$$\big[(F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\wedge\big(F^{\mathrm{Inv}}(\alpha_n)-\delta_n\big),\ (F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\vee\big(F^{\mathrm{Inv}}(\alpha_n)-\delta_n\big)\big],$$

respectively. If we can ensue that F is continuous on these intervals (with one-sided continuity on the boundaries), we can apply the extended Theorem of Rolle to further estimate (5.35) and (5.36).

By assumption, F is continuous in q_α and due to Assumption [C] also in a small neighbourhood $(q_\alpha-\Delta, q_\alpha+\Delta)$, $\Delta>0$. By (5.31) and since $\alpha_n\to\alpha$ we know that $(F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)$, $F^{\mathrm{Inv}}(\alpha_n)\in(q_\alpha-\Delta, q_\alpha+\Delta)$ for n big enough. Because ε_n, $\delta_n\to 0$ this then implies

$$\big[\big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\wedge F^{\mathrm{Inv}}(\alpha_n),\ \big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\vee F^{\mathrm{Inv}}(\alpha_n)\big]\subset(q_\alpha-\Delta, q_\alpha+\Delta)$$

and

$$\big[(F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\wedge\big(F^{\mathrm{Inv}}(\alpha_n)-\delta_n\big),\ (F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\vee\big(F^{\mathrm{Inv}}(\alpha_n)-\delta_n\big)\big]\subset(q_\alpha-\Delta, q_\alpha+\Delta)$$

if n is big enough, showing that F is indeed continuous on the desired intervals.

Thus, we are able to use the extended Theorem of Rolle on the inequalities (5.35) and (5.36). For the fraction in (5.35) the named theorem yields

$$\begin{aligned}\min\Big\{\big(\partial^-(F)(\xi_n)\big)^{-1}, \big(\partial^+(F)(\xi_n)\big)^{-1}\Big\}&\\ \le\left(\frac{F\big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)-F\big(F^{\mathrm{Inv}}(\alpha_n)\big)}{(F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n-F^{\mathrm{Inv}}(\alpha_n)}\right)^{-1}&\\ \le\max\Big\{\big(\partial^-(F)(\xi_n)\big)^{-1},\ \big(\partial^+(F)(\xi_n)\big)^{-1}\Big\}&\end{aligned}\tag{5.37}$$

for some $\xi_n\in\big(\big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\wedge F^{\mathrm{Inv}}(\alpha_n),\ \big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-\varepsilon_n\big)\vee F^{\mathrm{Inv}}(\alpha_n)\big)$. For the fraction in (5.36) the extended Theorem of Rolle gives

$$\begin{aligned}\min\Big\{\big(\partial^-(F)(\zeta_n)\big)^{-1}, \big(\partial^+(F)(\zeta_n)\big)^{-1}\Big\}&\\ \le\left(\frac{F\big((F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)\big)-F\big(F^{\mathrm{Inv}}(\alpha_n)-\delta_n\big)}{(F+t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n)-F^{\mathrm{Inv}}(\alpha_n)+\delta_n}\right)^{-1}&\\ \le\max\Big\{\big(\partial^-(F)(\zeta_n)\big)^{-1},\ \big(\partial^+(F)(\zeta_n)\big)^{-1}\Big\}&\end{aligned}\tag{5.38}$$

for a $\zeta_n \in \big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) \wedge \big(F^{\mathrm{Inv}}(\alpha_n) - \delta_n\big),\, (F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) \vee \big(F^{\mathrm{Inv}}(\alpha_n) - \delta_n\big)\big)$.

Next, we choose sequences $\alpha_n', \alpha_n'' \to \alpha$ with $F^{\mathrm{Inv}}(\alpha_n') = \xi_n$ and $F^{\mathrm{Inv}}(\alpha_n'') = \zeta_n$, which is possible due to the continuity of F^{Inv}. As F does not jump in ξ_n or ζ_n, the minima and maxima occurring in (5.37) and (5.38) can be rewritten using the hulls of $\partial^-(F^{\mathrm{Inv}})$, see i) above. We combine the observations so far with Lemma 2.13, i), Lemma 2.24 and (5.35) to obtain

$$\limsup_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \le \begin{cases} -\varphi(q_\alpha)\,\liminf_n \partial^-(F^{\mathrm{Inv}})_\wedge(\alpha_n') & \text{if } \varphi(q_\alpha) \ge 0,\\ -\varphi(q_\alpha)\,\limsup_n \partial^-(F^{\mathrm{Inv}})_\vee(\alpha_n') & \text{if } \varphi(q_\alpha) \le 0 \end{cases}$$

and

$$\liminf_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \ge \begin{cases} -\varphi(q_\alpha)\,\limsup_n \partial^-(F^{\mathrm{Inv}})_\vee(\alpha_n'') & \text{if } \varphi(q_\alpha) \ge 0,\\ -\varphi(q_\alpha)\,\liminf_n \partial^-(F^{\mathrm{Inv}})_\wedge(\alpha_n'') & \text{if } \varphi(q_\alpha) \le 0. \end{cases}$$

By the upper- and lower-semicontinuity of the hulls this finally gives

$$\limsup_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \le \begin{cases} -\varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha) & \text{if } \varphi(q_\alpha) \ge 0,\\ -\varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha) & \text{if } \varphi(q_\alpha) \le 0 \end{cases}$$

and

$$\liminf_n t_n^{-1}\Big(\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \ge \begin{cases} -\varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha) & \text{if } \varphi(q_\alpha) \ge 0,\\ -\varphi(q_\alpha)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha) & \text{if } \varphi(q_\alpha) \le 0. \end{cases}$$

This is the assertion in case 2), see the hulls in Corollary 5.23, ii).

Case 3) Assume $\alpha = F(q_\alpha-)$ for $q_\alpha \in \{y_1, \dots, y_r\}$. With (5.31) and the convergence $\alpha_n + \Delta_n \to \alpha$ we deduce $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) \le q_\alpha$ for n big enough, which shows that $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n \nearrow q_\alpha$ with $\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n \ne q_\alpha$ at least for big n. Additionally, since $\alpha_n \to \alpha$, it is certainly true that $F^{\mathrm{Inv}}(\alpha_n) \le q_\alpha$ again by choosing n big enough.

We first estimate the limes superior of (5.35), for which we partition the indizes n in the subsequence n_k with $F^{\mathrm{Inv}}(a_{n_k}) = q_\alpha$ and the remaining members. With the aid of Lemma 5.24 we are able to discuss the subsequences separately.

First, consider n_k as chosen before, assuming this subsequence has an infinite number of entries. To keep the notation simple, we assume without loss of generality that every index n fulfils $F^{\mathrm{Inv}}(a_n) = q_\alpha$. The fraction in (5.35) then becomes

$$\frac{\big(F + t_n\,\varphi_n\big)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n - q_\alpha}{F\big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n\big) - F(q_\alpha)},$$

which converges to 0, since the denominator tends to $F(q_\alpha-) - F(q_\alpha) \neq 0$ and the nominator goes to 0. Using this in (5.35) together with Lemma 2.13, i), implies

$$\limsup_n t_n^{-1}\Big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \leq 0$$

for the chosen subsequence.

Second, we consider the remaining indizes n_l of the original sequence n, so the ones with $F^{\mathrm{Inv}}(\alpha_{n_l}) < q_\alpha$. Again we assume without loss of generality that every n is chosen. Here, note that $\varphi_n((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n) \to \varphi(q_\alpha-)$ by the uniform convergence of φ_n and since it holds that $(F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - \varepsilon_n < q_\alpha$. In the present situation the fraction in (5.35) is the reciprocal of a difference quotient over an interval whose right endpoint is strictly smaller than q_α. Thus, F is continuous there and we can use the extended Theorem of Rolle as in case 2) together with Lemma 2.13, i), to obtain

$$\begin{aligned}
&\limsup_n t_n^{-1}\Big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \\
&\qquad\leq \begin{cases} -\varphi(q_\alpha-)\ \liminf_n \partial^-(F^{\mathrm{Inv}})_\wedge(\alpha'_n) & \text{if } \varphi(q_\alpha-) \geq 0, \\ -\varphi((q_\alpha-)\ \limsup_n \partial^-(F^{\mathrm{Inv}})_\vee(\alpha'_n) & \text{if } \varphi(q_\alpha-) \leq 0 \end{cases}
\end{aligned}$$

with $F^{\mathrm{Inv}}(\alpha'_n) = \xi_n$; note that $\alpha'_n < \alpha$ as ξ_n lies strictly below q_α. The properties of the semicontinuous hulls then imply

$$\limsup_n t_n^{-1}\Big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \leq \begin{cases} -\varphi(q_\alpha-)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha) & \text{if } \varphi(q_\alpha-) \geq 0, \\ -\varphi(q_\alpha-)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha) & \text{if } \varphi(q_\alpha-) \leq 0. \end{cases}$$

Combining the subsequences as in Lemma 5.24 yields

$$\begin{aligned}
&\limsup_n t_n^{-1}\Big((F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) - F^{\mathrm{Inv}}(\alpha_n)\Big) \\
&\qquad\leq \begin{cases} \Big(- \varphi(q_\alpha-)\,\partial^-(F^{\mathrm{Inv}})_\wedge(\alpha)\Big) \vee 0 & \text{if } \varphi(q_\alpha-) \geq 0, \\ \Big(- \varphi(q_\alpha-)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\Big) \vee 0 & \text{if } \varphi(q_\alpha-) \leq 0 \end{cases} \\
&\qquad= -\varphi(q_\alpha-)\,\partial^-(F^{\mathrm{Inv}})_\vee(\alpha)\,\mathbb{1}\left(\varphi(q_\alpha-) \leq 0\right),
\end{aligned}$$

which is the needed inequality in view of Corollary 5.23, iii).

For determining the limes inferior of (5.36) we observe that $F^{\mathrm{Inv}}(\alpha_n) \leq q_\alpha$ by (5.31). By separating the indizes n in the ones with $(F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) = q_\alpha$ and $(F + t_n\,\varphi_n)^{\mathrm{Inv}}(\alpha_n) <$

q_α and analogue arguments as before we can deduce

$$\liminf_n t_n^{-1}\Big((F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n)-F^{\text{Inv}}(\alpha_n)\Big)$$
$$\geq \begin{cases}\Big(-\varphi(q_\alpha-)\,\partial^-(F^{\text{Inv}})_\vee(\alpha)\Big)\wedge 0 & \text{if } \varphi(q_\alpha-)\geq 0,\\ \Big(-\varphi(q_\alpha-)\,\partial^-(F^{\text{Inv}})_\wedge(\alpha)\Big)\wedge 0 & \text{if } \varphi(q_\alpha-)\leq 0\end{cases}$$
$$= -\varphi(q_\alpha-)\,\partial^-(F^{\text{Inv}})_\vee(\alpha)\,\mathbb{1}\,(\varphi(q_\alpha-)\geq 0)\,.$$

This is the assertion for the limes inferior in case 3), see Corollary 5.23, iii).

Case 4) Here, $\alpha = F(q_\alpha)$ for some $q_\alpha \in \{y_1,\dots,<_r\}$. In this case, $(F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n)\geq q_\alpha$ is valid eventually and thus $\varphi_n\big((F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n)\big)\to\varphi(q_\alpha)$.

To estimate the limes superior of (5.35) we separate the occurring sequence according to $(F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n) = q_\alpha$ and $(F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n) > q_\alpha$ and discuss their greatest accumulation points. For the second subsequence, we possibly have to diminish ε_n to guarantee $(F+t_n\,\varphi_n)^{\text{Inv}}(\alpha_n)-\varepsilon_n > q_\alpha$ as well.

The limes inferior of (5.36) is estimated by examining the subsequences with $F^{\text{Inv}}(\alpha_n) = q_\alpha$ and $F^{\text{Inv}}(\alpha_n) > q_\alpha$; for the latter situation we potentially have to diminish δ_n to ensure $F^{\text{Inv}}(\alpha_n)-\delta_n > q_\alpha$ as well.

This finishes the proof of the theorem. □

5.8 Convergence in the Skorohod M_2-topology

In Section 2.3.3 we argued that the hypi-topology and the M_2-topology are equivalent in certain situations – in the present chapter we are in such a situation, as the sequence as well as the limit in Theorem 5.1 are (or can be chosen to be) càdlàg over a compact interval in $\mathbb{R}$.

In this section we show how to prove an M_2-version of Theorem 5.1, without using the equivalence of the topologies. Our goal is the following assertion.

5.25 Theorem.

Under the assumptions of Theorem 5.1, the standardized expectile process $\tau\mapsto\sqrt{n}(\hat{\mu}_{\tau,n}-\mu_\tau)$, $\tau\in[\tau_l,\tau_u]$, *converges weakly in* $\big(\ell^\infty([\tau_l,\tau_u]),\mathrm{d}_{s,2}\big)$ *to the limit process* $\big(\dot\psi_0^{\text{Inv}}(Z)(\tau)\big)_{\tau\in[\tau_l,\tau_u]}$, *where* $\dot\psi_0^{\text{Inv}}$ *and* $(Z_\tau)_{\tau\in[\tau_l,\tau_u]}$ *are as in Theorem 5.1.*

For the proof we require the same reduction steps for the M_2-distance $\mathrm{d}_{s,2}$ as in the proof of Theorem 5.1 for the hypi-distance d_{hypi}. However, working with the hypi-distance is easier since convergence is characterized by simple pointwise criteria, while the proof below is based on various results from Whitt (2002) for the M_2- and also the related M_1-distance.

Proof of Theorem 5.25. As indicated at the end of Step 2 in the proof of Theorem 5.1, the assertions in Steps 1 and 2 are independent of the topology used and thus apply here as well. Hence, we only show that (5.20), that is

$$\frac{\varphi_n}{c_{\iota_n}} \to \dot{\psi}_0^{\mathrm{Inv}}(\varphi) = \frac{\varphi}{c_0},$$

holds with respect to the M_2-distance as well. In order to reduce this to M_2-convergence of h_n to h (see (5.21)), we use Lemma 2.23 and the discussion thereafter to deal with fractions, products and sums, respectively.

It therefore remains to show convergence of h_n to h for the M_2-distance which can be done as follows. By Lemma 5.20 we have that $\varepsilon_n := \|\psi_0^{Inv}(t_n\varphi_n)(\cdot) - \mu_\cdot\| \to 0$ and using monotonicity of F it holds that

$$F(\mu_\tau - \varepsilon_n) \leq h_n(\tau) \leq F(\mu_\tau + \varepsilon_n)$$

for any $n \in \mathbb{N}$. If we can show that the upper and lower bound both converge to h with respect to $\mathrm{d}_{s,2}$, then Corollary 12.11.6, Whitt (2002), yields convergence of h_n to the same limit in the M_2-sense. We actually show convergence of the bounds to h with respect to the M_1-metric, which implies convergence in the M_2-sense; see Whitt (2002, Theorem 12.10.3). To this end, note that $\tau \mapsto F(\mu_\tau \pm \varepsilon_n)$ is an increasing function and $F(\mu_\tau \pm \varepsilon_n) \to F(\mu_\tau) = h(\tau)$ for every continuity point $\tau \in \mathcal{S}$, where $\mathcal{S}$ is defined in (5.22) and is dense. Thus, Corollary 12.5.1, Whitt (2002), can be applied to obtain the M_1- and hence M_2-convergence $F(\mu_\cdot \pm \varepsilon_n) \to h(\cdot)$. $\square$

5.9 Further technical proofs

Concluding the chapter, in this section we give the remaining details for the proofs in the former parts. This comprises the spared parts in the proof of Theorem 5.1, namely considering the increments and Lipschitz-continuity of $I_\tau(x;F)$ and basic but helpful bounds on $\tau + (1-2\tau)s$. Additionally, a representation for the increments of ψ_0^{Inv} and the remaining assertions leading to the consistency of the bootstrap in Theorem 5.4 are proven.

5.9.1 Proofs of Lemmas 5.17, 5.18 and 5.19

We start by calculating the increments of $I_\tau(x;F)$.

Proof of Lemma 5.17. Define the map $g(s) = I_\tau(x_2 + s(x_1 - x_2); F)$, $s \in [0,1]$, so that $I_\tau(x_1;F) - I_\tau(x_2;F) = g(1) - g(0)$. The map g is continuous, and in addition it is decreasing if $x_1 \geq x_2$, and increasing otherwise. Hence, it is of bounded variation and Theorem 7.23, Thomson et al. (2008), yields

$$g(1) - g(0) = \int_0^1 g'(s)\,\mathrm{d}s + \mu_g(\{s \in [0,1] \mid g'(s) = \pm\infty\}), \tag{5.39}$$

where μ_g is the Lebesgue-Stieltjes signed measure associated with g. From Holzmann and Klar (2016), the right- and left-sided derivatives of $I_\tau(x;F)$ are given by

$$\frac{\partial^+}{\partial x} I_\tau(x;F) = -(\tau + (1-2\,\tau)F(x)), \qquad \frac{\partial^-}{\partial x} I_\tau(x;F) = -(\tau + (1-2\,\tau)F(x-)).$$

Both derivatives are bounded by Lemma 5.18, such that $\{s \in [0,1] \mid g'(s) = \pm\infty\} = \emptyset$ in (5.39), and we obtain

$$\begin{aligned} I_\tau(x_1;F) - I_\tau(x_2;F) &= \int_0^1 g'(s)\,\mathrm{d}s \\ &= (x_1 - x_2) \int_0^1 -\big[\tau + (1-2\,\tau)\,F\big(x_2 + s(x_1 - x_2)\big)\big]\mathrm{d}s. \end{aligned}$$

□

Now we turn to the upper and lower bound of $\tau + (1-2\tau)s$.

Proof of Lemma 5.18. For the lower bound observe

$$\tau + (1-2\,\tau)s \begin{cases} = 1/2 & \text{if } \tau = 1/2 \\ \geq \tau & \text{if } \tau < 1/2 \\ \geq 1-\tau & \text{if } \tau > 1/2 \end{cases} \geq \min\{1/2, \tau_l, 1-\tau_u\} = \min\{\tau_l, 1-\tau_u\}.$$

The upper bound is proven similarly. □

The next remaining part in the proof of Theorem 5.1 is the Lipschitz-continuity of relevant maps.

Proof of Lemma 5.19. To show (5.14), which asserts the Lipschitz-continuity of $x \mapsto I_\tau(x;z)$, note that for $x_1 \leq x_2$

$$\begin{aligned} \big|I_\tau(x_1;z) - I_\tau(x_2;z)\big| &= \big|(x_2 - x_1)\,(\tau\mathbb{1}\,(z > x_1) + (1-\tau)\mathbb{1}\,(z \leq x_2)\,)\big| \\ &\leq |x_2 - x_1|. \end{aligned}$$

As for the Lipschitz-continuity of $\tau \mapsto I_\tau(x;z)$ in (5.15) observe

$$\begin{aligned} \big|I_\tau(x;z) - I_{\tau'}(x;z)\big| &= |\tau - \tau'|\big|(z-x)\mathbb{1}\,(z \geq x) + (x-z)\mathbb{1}\,(z < x)\,\big| \\ &\leq |\tau - \tau'|\,(|x| + |z|). \end{aligned}$$

For the Lipschitz-continuity of $\tau \mapsto \mu_\tau$, we use Corollary 1 of Beyn and Rieger (2011) for the function $x \mapsto I_\tau(x;F)$, $x \in B_R^{\mathbb{R}}(\mu_\tau)$, for appropriately chosen $R > 0$. We observe that

i) $x \mapsto I_\tau(x;F)$ is continuous for any $\tau \in [\tau_l, \tau_u]$, which is immediate from (5.14), and

ii) $x \mapsto I_\tau(x;F)$ fulfils

$$\left(I_\tau(x_1;F) - I_\tau(x_2;F)\right)(x_1 - x_2) \le -a\,(x_1 - x_2)^2$$

with $a = \min\{\tau_l, 1-\tau_u\} > 0$. This is clear from (5.12) and (5.13).

Let τ, $\tau' \in [\tau_l, \tau_u]$, and set $r = I_\tau(\mu_{\tau'};F)$. Using i) above yields

$$\frac{1}{a}\,|r| = \frac{1}{a}\,|I_\tau(\mu_{\tau'};F) - I_\tau(\mu_\tau;F)| \le |\mu_\tau - \mu_{\tau'}| \le \mu_{\tau_u} - \mu_{\tau_l}.$$

Choosing $R = \mu_{\tau_u} - \mu_{\tau_l} + \varepsilon$ for some small $\varepsilon > 0$ yields the inclusions $[\mu_{\tau_l}, \mu_{\tau_u}] \subset B_R^{\mathbb{R}}(\mu_\tau) \subset [\mu_{\tau_l} - R, \mu_{\tau_u} + R]$, $\tau \in [\tau_l, \tau_u]$, since $\mu_\tau \in [\mu_{\tau_l}, \mu_{\tau_u}]$. Corollary 1, Beyn and Rieger (2011), now gives an $\bar{x} \in B_R^{\mathbb{R}}(\mu_\tau)$ with $I_\tau(\bar{x};F) = r$ and

$$|\mu_\tau - \bar{x}| \le \frac{1}{a}\,|r|.$$

Since $x \mapsto I_\tau(x;F)$ is strictly decreasing and $[\mu_{\tau_l}, \mu_{\tau_u}] \subset B_R^{\mathbb{R}}(\mu_\tau)$ as well as $I_\tau(\bar{x};F) = r = I_\tau(\mu_{\tau'};F)$, we obtain $\bar{x} = \mu_{\tau'}$. We conclude that

$$\begin{aligned}
|\mu_\tau - \mu_{\tau'}| &\le \frac{1}{a}\,|I_\tau(\mu_{\tau'};F)| = \frac{1}{a}\,\left|\mathbb{E}\left[I_\tau(\mu_{\tau'},Y) - I_{\tau'}(\mu_{\tau'},Y)\right]\right| \\
&\le |\tau - \tau'|\,\frac{|\mu_{\tau'}| + \mathbb{E}\left[|Y|\right]}{a} \le |\tau - \tau'|\,\frac{|\mu_{\tau_u}| \vee |\mu_{\tau_l}| + \mathbb{E}\left[|Y|\right]}{a},
\end{aligned} \tag{5.40}$$

where we used (5.15). □

5.9.2 Shape of increments of ψ_0^{Inv}

For proving the representation of $\psi_0^{\mathrm{Inv}}(t\,\nu) - \psi_0^{\mathrm{Inv}}(0)$ of Lemma 5.20, we use the increments of $I_\tau(x;F)$.

Proof of Lemma 5.20. For the first statement, given $\rho \in \ell^\infty([\tau_l, \tau_u])$, it follows from (5.12) that

$$\begin{aligned}
\psi_0(\rho)(\tau) &= \psi_0(\mu_\cdot + (\rho - \mu_\cdot))(\tau) - \psi_0(\mu_\cdot)(\tau) \\
&= -\Big(I_\tau(\mu_\tau + (\rho(\tau) - \mu_\tau);F) - I_\tau(\mu_\tau;F)\Big) \\
&= (\rho(\tau) - \mu_\tau)\,\Big[\tau + (1 - 2\,\tau)\int_0^1 F(\mu_\tau + s(\rho(\tau) - \mu_\tau))\,\mathrm{d}s\Big].
\end{aligned}$$

The term in angle brackets on the right hand side is bounded away from zero by (5.13) and thus choosing $\rho = \psi_0^{\mathrm{Inv}}(t\,\nu)$, observing $\mu_\tau = \psi_0^{\mathrm{Inv}}(0)(\tau)$, and reorganising the above equation leads to (5.17) as

$$\begin{aligned}
&t^{-1}\Big(\psi_0^{\mathrm{Inv}}(0 + t\,\nu) - \psi_0^{\mathrm{Inv}}(0)\Big)(\tau)\\
&= t^{-1}\psi_0\big(\psi_0^{\mathrm{Inv}}(t\,\nu)\big)(\tau)\left[\tau + (1-2\,\tau)\int_0^1 F\big(\mu_\tau + s(\psi_0^{\mathrm{Inv}}(t\,\nu)(\tau) - \mu_\tau)\big)\,\mathrm{d}s\right]^{-1}\\
&= \nu(\tau)\left[\tau + (1-2\,\tau)\int_0^1 F\big(\mu_\tau + s(\psi_0^{\mathrm{Inv}}(t\,\nu)(\tau) - \mu_\tau)\big)\,\mathrm{d}s\right]^{-1}.
\end{aligned}$$

Now, for the second part, remembering (5.19) and (5.13), $\min\{\tau_l, 1-\tau_u\} \le c_\varphi \le 3/2$ holds uniformly for any φ. Set $\varphi_n(\cdot) = \big(\psi_0^{\mathrm{Inv}}(\nu_n)(\cdot) - \mu.\big)$, then (5.17) yields with $t = 1$

$$\big\|\psi_0^{\mathrm{Inv}}(\nu_n)(\cdot) - \mu.\big\| \le \|\nu_n\|\,\|c_{\varphi_n}\|^{-1} \le \|\nu_n\|\big(\min\{\tau_l, 1-\tau_u\}\big)^{-1} \to 0.$$

Last, (5.18) follows by continuity of $\tau \mapsto \mu_\tau$ (see Lemma 5.19). □

5.9.3 Step 1 for the bootstrap

We end the chapter by showing the remaining parts in the proof of the bootstrap consistency result, Theorem 5.4, which are parallel to Step 1 in the proof of Theorem 5.1.

Proof of Lemma 5.21. First consider (5.23). We start with individual consistency, the proof of which is inspired by Lemma 5.10, van der Vaart (1998).

Since $I_\tau(\mu^*_{\tau,n}; F^*_n) = 0$ and $x \mapsto I_\tau(x; F^*_n)$ is strictly decreasing, for any $\varepsilon, \eta_l, \eta_u > 0$ the inequality $I_\tau(\widehat{\mu}_{\tau,n} - \varepsilon; F^*_n) > \eta_l$ implies $\mu^*_{\tau,n} > \widehat{\mu}_{\tau,n} - \varepsilon$ and from $I_\tau(\widehat{\mu}_{\tau,n} + \varepsilon; F^*_n) < -\eta_u$ it follows that $\mu^*_{\tau,n} < \widehat{\mu}_{\tau,n} + \varepsilon$. Thus,

$$\mathrm{P}^*_n\big(I_\tau(\widehat{\mu}_{\tau,n} - \varepsilon; F^*_n) > \eta_l,\, I_\tau(\widehat{\mu}_{\tau,n} + \varepsilon; F^*_n) < -\eta_u\big) \le \mathrm{P}^*_n\big(\widehat{\mu}_{\tau,n} - \varepsilon < \mu^*_{\tau,n} < \widehat{\mu}_{\tau,n} + \varepsilon\big)$$

holds, and it suffices to show almost sure convergence of the left hand side to 1 for appropriately chosen $\eta_l, \eta_u > 0$, for which it is enough to deduce the almost sure convergence $P^*_n\big(I_\tau(\widehat{\mu}_{\tau,n} - \varepsilon; F^*_n) > \eta_l\big) \to 1$ and $P^*_n\big(I_\tau(\widehat{\mu}_{\tau,n} + \varepsilon; F^*_n) < -\eta_u\big) \to 1$. Choose $2\,\eta_l = I_\tau(\mu_\tau - \varepsilon; F) \neq 0$ to obtain the estimate

$$\begin{aligned}
\mathrm{P}^*_n\big(I_\tau(\widehat{\mu}_{\tau,n} - \varepsilon; F^*_n) > \eta_l\big) &\ge \mathrm{P}^*_n\big(\big|I_\tau(\mu_\tau - \varepsilon; F) - |I_\tau(\hat{\mu}_{\tau,n} - \varepsilon; F^*_n) - I_\tau(\mu_\tau - \varepsilon; F)|\big| > \eta_l\big)\\
&\ge \mathrm{P}^*_n\big(|I_\tau(\hat{\mu}_{\tau,n} - \varepsilon; F^*_n) - I_\tau(\mu_\tau - \varepsilon; F)| < \eta_l\big)
\end{aligned}$$

with the inverse triangle inequality. Similarly, for $2\,\eta_u = I_\tau(\mu_\tau + \varepsilon; F) \neq 0$

$$\mathrm{P}^*_n\big(I_\tau(\widehat{\mu}_{\tau,n} + \varepsilon; F^*_n) < -\eta_u\big) \ge \mathrm{P}^*_n\big(|I_\tau(\widehat{\mu}_{\tau,n} + \varepsilon; F^*_n) - I_\tau(\mu_\tau + \varepsilon; F)| < \eta_u\big)$$

is true. In both inequalities the right hand side converges to 1 almost surely, provided that almost surely, $I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon; F_n^*) \to I_\tau(\mu_\tau \pm \varepsilon; F)$ in probability conditionally on $Y_1, Y_2, \ldots$. To this end, start with

$$\mathbb{E}_n^* \left[I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon; F_n^*) \right] = \sum_{i=1}^n \mathrm{P}_n^*(Y_1^* = Y_i)\, I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon, Y_i) = I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon; F_n),$$

so that it remains to show convergence of the right hand side to $I_\tau(\mu_\tau \pm \varepsilon; F)$ for almost every sequence $Y_1, Y_2, \ldots$. For this purpose, use Lemma 5.19 to deduce

$$\left| I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon; F_n) - I_\tau(\mu_\tau \pm \varepsilon; F_n) \right| \leq \left| \widehat{\mu}_{\tau,n} - \mu_\tau \right|,$$

where the upper bound converges to 0 by the strong consistency of $\widehat{\mu}_{\tau,n}$ (Holzmann and Klar, 2016, Theorem 2). Further, $I_\tau(\mu_\tau \pm \varepsilon; F_n) \to I_\tau(\mu_\tau \pm \varepsilon; F)$ almost surely by the strong law of large numbers, thus $I_\tau(\widehat{\mu}_{\tau,n} \pm \varepsilon; F_n)$ converges to $I_\tau(\mu_\tau \pm \varepsilon; F)$ for almost every sequence $Y_1, Y_2, \ldots$, what concludes the proof of individual consistency of $\mu_{\tau,n}^*$.

To strengthen this to uniform consistency, we use a Glivenko-Cantelli argument as in Holzmann and Klar (2016), Theorem 2. Let $d_n = \widehat{\mu}_{\tau_u,n} - \widehat{\mu}_{\tau_l,n}$ and observe that $d_n \to d = \mu_{\tau_u} - \mu_{\tau_l}$ almost surely by the strong consistency of $\widehat{\mu}_{\tau,n}$. Let $r \in \mathbb{N}$ and choose $\tau_l = \tau_0 \leq \tau_1 \leq \ldots \leq \tau_r = \tau_u$ such that

$$\widehat{\mu}_{\tau_s,n} = \widehat{\mu}_{\tau_l,n} + \frac{s\, d_n}{r}$$

for every $s \in \{1, \ldots, r\}$, which is possible because of the continuity of $\tau \mapsto \widehat{\mu}_{\tau,n}$. As the expectile functional is strictly increasing in τ it follows that

$$\mu_{\tau_s,n}^* - \widehat{\mu}_{\tau_{s+1},n} \leq \mu_{\tau,n}^* - \widehat{\mu}_{\tau,n} \leq \mu_{\tau_{s+1},n}^* - \widehat{\mu}_{\tau_s,n}$$

for $\tau_s \leq \tau \leq \tau_{s+1}$. This implies

$$\left\| \mu_{\tau,n}^* - \widehat{\mu}_{\tau,n} \right\| \leq \max_{1 \leq s \leq r} \left| \mu_{\tau_s,n}^* - \widehat{\mu}_{\tau_s,n} \right| + \frac{d_n}{r},$$

hence

$$\limsup_n \left\| \mu_{\tau,n}^* - \widehat{\mu}_{\tau,n} \right\| \leq \limsup_n \max_{1 \leq s \leq r} \left| \mu_{\tau_s,n}^* - \widehat{\mu}_{\tau_s,n} \right| + \limsup_n \frac{d_n}{r} = \frac{d}{r}$$

holds conditionally in probability for almost every sequence $Y_1, Y_2, \ldots$. Letting $r \to \infty$ completes the proof of strong bootstrap consistency in (5.23).

Let us turn to (5.24). The idea is to use Theorem 19.28, van der Vaart (1998), for the random class

$$\mathcal{H}_n = \left\{ z \mapsto -I_\tau(\widehat{\mu}_{\tau,n}; z) \;\middle|\; \tau \in [\tau_l, \tau_u] \right\},$$

which is a subset of

$$\mathcal{H}_{\eta_n} = \{z \mapsto -I_\tau(\mu_\tau + x; z) \mid |x| \leq \eta_n, \, \tau \in [\tau_l, \tau_u]\}$$

for the sequence $\eta_n = \|\widehat{\mu}_{\tau,n} - \mu_\tau\|$. Hence, almost surely it holds that

$$J_{[\,]}(\varepsilon, \mathcal{H}_n, \|\cdot\|_{n,2}) \leq J_{[\,]}(\varepsilon, \mathcal{H}_{\eta_n}, \|\cdot\|_{n,2}).$$

The class $\mathcal{H}_{\eta_n}$ has envelope $(|\mu_{\tau_l}| + |\mu_{\tau_u}| + \eta_n + z)$, which satisfies the Lindeberg condition. By (5.16) the class $\mathcal{H}_{\eta_n}$ is a class consisting of Lipschitz-functions with Lipschitz-constant given by $L_n(z) = (C + \eta_n + |z|)$ for some $C \geq 1$, such that the bracketing number fulfils

$$N_{[\,]}(\delta, \mathcal{H}_{\eta_n}, \|\cdot\|_{n,2}) \leq C_1 \left[\mathbb{E}_n \left[L_n(Y)^2\right] \frac{\eta_n + \tau_u - \tau_l}{\delta}\right]^2$$

by Example 19.7, van der Vaart (1998), where C_1 is some constant not depending on n and $\|h\|_{n,2} = \mathbb{E}_n[h(Y)^2]^{1/2}$. By the strong law of large numbers, $\mathbb{E}_n\left[L_n(Y)^2\right] \to \mathbb{E}\left[(C + |Y|)^2\right]$ holds almost surely, in addition $\eta_n \to 0$ almost surely by the strong consistency of $\widehat{\mu}_{\tau,n}$, such that the above bracketing number is of order δ^{-2}. Thus, the bracketing integral $J_{[\,]}(\varepsilon_n, \mathcal{H}_{\eta_n}, \|\cdot\|_{n,2})$ converges to 0 almost surely for every sequence $\varepsilon_n \searrow 0$.

Next, we show the almost sure convergence of $\mathbb{E}_n\left[I_\tau(\widehat{\mu}_{\tau,n}; Y) I_{\tau'}(\widehat{\mu}_{\tau',n}; Y)\right]$ to the limit $\mathbb{E}\left[I_\tau(\mu_\tau; Y) I_{\tau'}(\mu_{\tau'}; Y)\right]$. First, it holds that

$$\begin{aligned}
\mathbb{E}_n\left[I_\tau(\widehat{\mu}_{\tau,n}; Y) I_{\tau'}(\widehat{\mu}_{\tau',n}; Y)\right] &= \mathbb{E}_n\left[I_\tau(\widehat{\mu}_{\tau,n}; Y)\left(I_{\tau'}(\widehat{\mu}_{\tau',n}; Y) - I_{\tau'}(\mu_{\tau'}; Y)\right)\right] \\
&\quad + \mathbb{E}_n\left[I_{\tau'}(\mu_{\tau'}; Y)\left(I_\tau(\widehat{\mu}_{\tau,n}; Y) - I_\tau(\mu_\tau; Y)\right)\right] \\
&\quad + \mathbb{E}_n\left[I_\tau(\mu_\tau; Y) I_{\tau'}(\mu_{\tau'}; Y)\right],
\end{aligned}$$

where the last summand converges almost surely to $\mathbb{E}\left[I_\tau(\mu_\tau; Y) I_{\tau'}(\mu_{\tau'}; Y)\right]$ by the strong law of large numbers. For the first term we estimate

$$\mathbb{E}_n\left[|I_\tau(\widehat{\mu}_{\tau,n}; Y)|\,|I_{\tau'}(\widehat{\mu}_{\tau',n}; Y) - I_{\tau'}(\mu_{\tau'}; Y)|\right] \leq \left[|\widehat{\mu}_{\tau,n}| + \mathbb{E}_n\left[|Y|\right]\right] |\widehat{\mu}_{\tau',n} - \mu_{\tau'}|$$

with the aid of Lemma 5.19, where the upper bound converges to 0 almost surely by the strong law of large numbers, strong consistency of $\widehat{\mu}_{\tau',n}$ and boundedness of $\widehat{\mu}_{\tau,n}$, which in fact also follows from the strong consistency of the empirical expectile and since $\tau \in [\tau_l, \tau_u]$. The remaining summand above is treated likewise, hence, the sequence of expectations $\mathbb{E}_n\left[I_\tau(\widehat{\mu}_{\tau,n}; Y) I_{\tau'}(\widehat{\mu}_{\tau',n}; Y)\right]$ indeed converges almost surely to the stated limit. The assertion (5.24) now follows from Theorem 19.28, van der Vaart (1998).

Now we show (5.25). Setting $\eta_n = \|\widehat{\mu}_{\tau,n} - \mu_\tau\|$ again, as a first step we can estimate

$$\begin{aligned}
&\sup_{\|\varphi\| \leq \delta_n} \left\|\psi_n^*(\widehat{\mu}_{\cdot,n} + \varphi)(\cdot) - \psi_n(\widehat{\mu}_{\cdot,n} + \varphi)(\cdot) - \left[\psi_n^*(\widehat{\mu}_{\cdot,n})(\cdot) - \psi_n(\widehat{\mu}_{\cdot,n})(\cdot)\right]\right\| \\
&= \sup_{\|\varphi\| \leq \delta_n} \sqrt{n}\left\|\psi_n^*(\mu_\cdot + (\widehat{\mu}_{\cdot,n} - \mu_\cdot + \varphi))(\cdot) - \psi_n(\mu_\cdot + (\widehat{\mu}_{\cdot,n} - \mu_\cdot + \varphi))(\cdot)\right. \\
&\qquad\qquad \left. - \left[\psi_n^*(\mu_\cdot + (\widehat{\mu}_{\cdot,n} - \mu_\cdot))(\cdot) - \psi_n(\mu_\cdot + (\widehat{\mu}_{\cdot,n} - \mu_\cdot))(\cdot)\right]\right\| \\
&\leq \sup_{|x_1|, |x_2| \leq \delta_n + \eta_n} \sqrt{n}\left\|\psi_n^*(\mu_\cdot + x_1)(\cdot) - \psi_n(\mu_\cdot + x_1)(\cdot) - \left[\psi_n^*(\mu_\cdot + x_2)(\cdot) - \psi_n(\mu_\cdot + x_2)(\cdot)\right]\right\|,
\end{aligned}$$

such that for (5.25) it suffices to show almost sure convergence of the conditional expectations $\mathbb{E}_n^*\left[\|\sqrt{n}(\mathbb{E}_n^* - \mathbb{E}_n)\|_{\mathcal{H}_{\nu_n}}\right]$ to 0 for the class

$$\mathcal{H}_{\nu_n} = \{z \mapsto I_\tau(\mu_\tau + x_1; z) - I_\tau(\mu_\tau + x_2; z) \mid |x_1|, |x_2| \le \nu_n,\, \tau \in [\tau_l, \tau_u]\},$$

where $\nu_n = \delta_n + \eta_n$. We use Corollary 19.35, van der Vaart (1998), which implies that

$$\mathbb{E}_n^*\left[\|\sqrt{n}(\mathbb{E}_n^* - \mathbb{E}_n)\|_{\mathcal{H}_{\nu_n}}\right] \le C_2 J_{[\,]}\Big(\mathbb{E}_n\left[L_n(Y)^2\right], \mathcal{H}_{\nu_n}, \|\cdot\|_{n,2}\Big)$$

almost surely, where $L_n(z)$ is an envelope function for $\mathcal{H}_{\nu_n}$ and C_2 is some constant. The proof of the convergence of the conditional expectations above consists of finding this envelope and determining the order of the bracketing integral. By Lemma 5.19 every function in the class $\mathcal{H}_{\nu_n}$ is Lipschitz-continuous, as for any $\tau, \tau' \in [\tau_l, \tau_u]$ as well as $x_1, x_1', x_2, x_2' \in [-\nu_n, \nu_n]$ the almost sure inequality

$$\begin{aligned} &\big|I_\tau(\mu_\tau + x_1; z) - I_\tau(\mu_\tau + x_2; z) - \big(I_{\tau'}(\mu_{\tau'} + x_1'; z) - I_{\tau'}(\mu_{\tau'} + x_2'; z)\big)\big| \\ &\quad \le \left(|x_1 - x_1'| + |x_2 - x_2'| + |\tau - \tau'|\right) 2\left(C_3 + \nu_n + |z|\right) \end{aligned} \tag{5.41}$$

is true for some constant $C_3 > 0$; see also (5.16). Thus, using the same arguments as above, the bracketing integral $J_{[\,]}(\varepsilon_n, \mathcal{H}_{\nu_n}, \|\cdot\|_{n,2})$ converges to 0 almost surely for any sequence $\varepsilon_n \searrow 0$. Finally, by (5.41) the function $L_n(z) = \eta_n 4\left(C_3 + |z|\right)$ is an envelope for $\mathcal{H}_{\nu_n}$. By the strong law of large numbers, the square integrability of Y and since $\eta_n \to 0$ almost surely, it holds that $\mathbb{E}_n\left[L_n(Y)^2\right] \to 0$ almost surely, so that

$$\mathbb{E}_n^*\left[\|\sqrt{n}(\mathbb{E}_n^* - \mathbb{E}_n)\|_{\mathcal{H}_{\nu_n}}\right] \le J_{[\,]}\Big(\mathbb{E}_n\left[L_n(Y)^2\right], \mathcal{H}_{\nu_n}, \|\cdot\|_{n,2}\Big) \to 0$$

is valid for almost every sequence $Y_1, Y_2, \ldots$, which concludes the proof of (5.25) by utilizing Markov's inequality.

It remains to establish (5.26). By (5.23) and (5.25), and since $\psi_n(\widehat{\mu}_{\cdot,n}), \psi_n^*(\mu_{\cdot,n}^*) = 0$, we have that almost surely, conditionally on $Y_1, Y_2, \ldots$,

$$\begin{aligned} \sqrt{n}\big(\psi_n(\mu_{\cdot,n}^*) - \psi_n(\widehat{\mu}_{\cdot,n})\big) &= \sqrt{n}\big(\psi_n(\mu_{\cdot,n}^*) - \psi_n^*(\mu_{\cdot,n}^*)\big) \\ &= -\sqrt{n}\big(\psi_n^*(\widehat{\mu}_{\cdot,n}) - \psi_n(\widehat{\mu}_{\cdot,n})\big) + o_{\mathrm{P}_n^*}(1) \end{aligned}$$

is valid, where we used similar arguments as those which led to (5.6). The right hand side converges to $-Z$ conditionally on $Y_1, Y_2, \ldots$ in distribution, almost surely, by (5.24), which equals Z in distribution. □

Bibliography

Acerbi, C., C. Nordio, and C. Sirtori (2001). "Expected Shortfall as a Tool for Financial Risk Management". In: *Journal of Emerging Market Finance* 8.3, pp. 87–107.

Acerbi, C. and D. Tasche (2002a). "Expected Shortfall: A Natural Coherent Alternative to Value at Risk". In: *Economic Notes* 31.2, pp. 379–388. ISSN: 1468-0300. DOI: 10.1111/1468-0300.00091. URL: http://dx.doi.org/10.1111/1468-0300.00091.

Acerbi, C. and D. Tasche (2002b). "On the coherence of expected shortfall". In: *Journal of Banking and Finance* 26.7, pp. 1487–1503. ISSN: 0378-4266. DOI: https://doi.org/10.1016/S0378-4266(02)00283-2.

Agarwal, A. and S. Agarwal (July 2015). "On Consistent Surrogate Risk Minimization and Property Elicitation". In: *Proceedings of The 28th Conference on Learning Theory*. Ed. by P. Grünwald, E. Hazan, and S. Kale. Vol. 40. Proceedings of Machine Learning Research. Paris, France: PMLR, pp. 4–22. URL: http://proceedings.mlr.press/v40/Agarwal15.html.

Andrews, D. and D. Pollard (1994). "An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes". In: *International Statistical Review / Revue Internationale de Statistique* 62.1, pp. 119–132.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). "Coherent Measures of Risk". In: *Mathematical Finance* 9.3, pp. 203–228. ISSN: 1467-9965. DOI: 10.1111/1467-9965.00068. URL: http://dx.doi.org/10.1111/1467-9965.00068.

Attouch, H. and R.J.-B. Wets (Oct. 1980). *Approximation and Convergence in Nonlinear Optimization.* Working Paper. Laxenburg, Austria: IIASA. URL: http://pure.iiasa.ac.at/id/eprint/1327/.

Avram, F. and M.S. Taqqu (1989). "Probability Bounds for M-Skorohod Oscillations". In: *Stochastic Processes and their Applications* 33.1, pp. 63–72. ISSN: 0304-4149. DOI: https://doi.org/10.1016/0304-4149(89)90066-5. URL: http://www.sciencedirect.com/science/article/pii/0304414989900665.

Bahadur, R.R. (1966). "A note on quantiles in large samples." In: *Ann. Math. Stat.* 37, pp. 577–580. ISSN: 0003-4851. DOI: 10.1214/aoms/1177699450.

Basel (2013). *Fundamental review of the trading book: A revised market risk framework. https://www.bis.org/publ/bcbs265.pdf.*

Basel (2017). *Basel III: A global regulatory framework for more resilient banks and banking systems. https://www.bis.org/publ/bcbs189.pdf.*

Bellini, F. and V. Bignozzi (2015). “On elicitable risk measures”. In: *Quantitative Finance* 15.5, pp. 725–733. DOI: 10.1080/14697688.2014.946955. eprint: https://doi.org/10.1080/14697688.2014.946955. URL: https://doi.org/10.1080/14697688.2014.946955.

Bellini, F., B. Klar, A. Müller, and E.R. Gianin (2014). “Generalized quantiles as risk measures.” In: *Insur. Math. Econ.* 54, pp. 41–48. ISSN: 0167-6687. DOI: 10.1016/j.insmatheco.2013.10.015.

Beyn, W.-J. and J. Rieger (2011). “An Implicit Function Theorem for One-sided Lipschitz Mappings”. In: *Set-Valued and Variational Analysis* 19.3, pp. 343–359. ISSN: 1877-0541. DOI: 10.1007/s11228-010-0162-8. URL: http://dx.doi.org/10.1007/s11228-010-0162-8.

Billingsley, P. (1999). *Convergence of Probability Measures*. 2nd ed. Wiley Series in Probability and Statistics. Wiley. ISBN: 9781118625965.

Bücher, A. and I. Kojadinovic (Mar. 2018). “A Note on Conditional Versus Joint Unconditional Weak Convergence in Bootstrap Consistency Results”. In: *Journal of Theoretical Probability*, pp. 1–21. ISSN: 1572-9230. DOI: 10.1007/s10959-018-0823-3. URL: https://doi.org/10.1007/s10959-018-0823-3.

Bücher, A., J. Segers, and S. Volgushev (2014). “When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs.” In: *Ann. Stat.* 42.4, pp. 1598–1634. ISSN: 0090-5364; 2168-8966/e. DOI: 10.1214/14-AOS1237.

Cai, Z. and X. Wang (2008). “Nonparametric estimation of conditional VaR and expected shortfall”. In: *Journal of Econometrics* 147.1. Econometric modelling in finance and risk management: An overview, pp. 120–130. ISSN: 0304-4076. DOI: http://dx.doi.org/10.1016/j.jeconom.2008.09.005. URL: http://www.sciencedirect.com/science/article/pii/S0304407608001292.

Chen, S.X. (2008). “Nonparametric Estimation of Expected Shortfall”. In: *Journal of Financial Econometrics* 6.1, pp. 87–107. DOI: 10.1093/jjfinec/nbm019. eprint: http://jfec.oxfordjournals.org/content/6/1/87.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/6/1/87.abstract.

Chen, S.X. and C.Y. Tang (2005). “Nonparametric Inference of Value-at-Risk for Dependent Financial Returns”. In: *Journal of Financial Econometrics* 3.2, pp. 227–255. DOI: 10.1093/jjfinec/nbi012. eprint: http://jfec.oxfordjournals.org/content/3/2/227.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/3/2/227.abstract.

Chibisov, D.M. (1965). “An Investigation of the Asymptotic Power of the Tests of Fit”. In: *Theory of Probability & Its Applications* 10.3, pp. 421–437. DOI: 10.1137/1110050. eprint: https://doi.org/10.1137/1110050. URL: https://doi.org/10.1137/1110050.

Clark, T.E. and M.W. McCracken (2001). “Tests of equal forecast accuracy and encompassing for nested models”. In: *Journal of Econometrics* 105.1. Forecasting and empirical methods in finance and macroeconomics, pp. 85–110. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(01)00071-9. URL: http://www.sciencedirect.com/science/article/pii/S0304407601000719.

Danielsson, J., B. Jorgensen, S. Mandira, G. Samorodnitsky, and C. de Vries (2005). *Subadditivity re–examined: the case for value-at-risk*. LSE Research Online Documents on Economics. London School of Economics and Political Science, LSE Library. URL: https://EconPapers.repec.org/RePEc:ehl:lserod:24668.

Davis, R.A., K. Knight, and J. Liu (1992). "M-estimation for autoregressions with infinite variance". In: *Stochastic Processes and their Applications* 40.1, pp. 145–180. ISSN: 0304–4149. DOI: http://dx.doi.org/10.1016/0304-4149(92)90142-D.

Dehling, H., O. Durieu, and D. Volny (2009). "New techniques for empirical processes of dependent data". In: *Stochastic Processes and their Applications* 119.10, pp. 3699–3718. DOI: https://doi.org/10.1016/j.spa.2009.07.003.

Dehling, H., T. Mikosch, and M. Sörensen (2012). *Empirical Process Techniques for Dependent Data*. SpringerLink : Bücher. Birkhäuser Boston.

Diebold, F.X. and R.S. Mariano (1995). "Comparing Predictive Accuracy". In: *Journal of Business & Economic Statistics* 13.3, pp. 253–265. DOI: 10.1198/073500102753410444.

Ehm, W., T. Gneiting, A. Jordan, and F. Krüger (2016). "Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78.3, pp. 505–562. ISSN: 1467-9868. DOI: 10.1111/rssb.12154. URL: http://dx.doi.org/10.1111/rssb.12154.

Embrechts P. and Hofert, M. (2013). "A note on generalized inverses". In: *Mathematical Methods of Operations Research* 77.3, pp. 423–432. ISSN: 1432-5217. DOI: 10.1007/s00186-013-0436-7. URL: https://doi.org/10.1007/s00186-013-0436-7.

Emmer, S., M. Kratz, and D. Tasche (2015). "What is the best risk measure in practice? A comparison of standard measures". In: *Journal of Risk* 18.2, pp. 31–60. DOI: 10.21314/JOR.2015.318.

Fissler, T. and J.F. Ziegel (Aug. 2016). "Higher order elicitability and Osband's principle". In: *Ann. Statist.* 44.4, pp. 1680–1707. DOI: 10.1214/16-AOS1439. URL: https://doi.org/10.1214/16-AOS1439.

Frongillo, R. and I.A. Kash (2015a). "On Elicitation Complexity". In: *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. NIPS'15. Montreal, Canada: MIT Press, pp. 3258–3266. URL: http://dl.acm.org/citation.cfm?id=2969442.2969603.

Frongillo, R. and I.A. Kash (July 2015b). "Vector-Valued Property Elicitation". In: *Proceedings of The 28th Conference on Learning Theory*. Ed. by P. Grünwald, E.H., and S. Kale. Vol. 40. Proceedings of Machine Learning Research. Paris, France: PMLR, pp. 710–727. URL: http://proceedings.mlr.press/v40/Frongillo15.html.

Gneiting, T. (2011a). “Making and Evaluating Point Forecasts”. In: *Journal of the American Statistical Association* 106.494, pp. 746–762. DOI: 10.1198/jasa.2011.r10138. eprint: http://dx.doi.org/10.1198/jasa.2011.r10138.

Gneiting, T. (2011b). “Quantiles as optimal point forecasts”. In: *International Journal of Forecasting* 27.2, pp. 197–207. ISSN: 0169-2070. DOI: https://doi.org/10.1016/j.ijforecast.2009.12.015. URL: http://www.sciencedirect.com/science/article/pii/S0169207010000063.

Gneiting, T. and A.E. Raftery (2007). “Strictly Proper Scoring Rules, Prediction, and Estimation”. In: *Journal of the American Statistical Association* 102.477, pp. 359–378. DOI: 10.1198/016214506000001437. eprint: https://doi.org/10.1198/016214506000001437. URL: https://doi.org/10.1198/016214506000001437.

Heinrich, C. (2014). “The mode functional is not elicitable”. In: *Biometrika* 101.1, pp. 245–251. URL: https://EconPapers.repec.org/RePEc:oup:biomet:v:101:y:2014:i:1:p:245-251..

Hill, J.B. (2013). “Expected Shortfall Estimation and Gaussian Inference for Infinite Variance Time Series”. In: *Journal of Financial Econometrics* 13.1, pp. 1–44. DOI: 10.1093/jjfinec/nbt020. eprint: http://jfec.oxfordjournals.org/content/early/2013/09/13/jjfinec.nbt020.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/early/2013/09/13/jjfinec.nbt020.abstract.

Holzmann, H. and B. Klar (2016). “Expectile asymptotics”. In: *Electron. J. Statist.* 10.2, pp. 2355–2371. DOI: 10.1214/16-EJS1173. URL: http://dx.doi.org/10.1214/16-EJS1173.

Huber, P.J. (Mar. 1964). “Robust Estimation of a Location Parameter”. In: *Ann. Math. Statist.* 35.1, pp. 73–101. DOI: 10.1214/aoms/1177703732. URL: https://doi.org/10.1214/aoms/1177703732.

Huber, P.J. (1967). “The behavior of maximum likelihood estimates under nonstandard conditions”. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics.* Berkeley, Calif.: University of California Press, pp. 221–233. URL: https://projecteuclid.org/euclid.bsmsp/1200512988.

Jones, M.C. (1994). “Expectiles and M-quantiles are quantiles”. In: *Statistics & Probability Letters* 20.2, pp. 149–153. ISSN: 0167-7152. DOI: https://doi.org/10.1016/0167-7152(94)90031-0. URL: http://www.sciencedirect.com/science/article/pii/0167715294900310.

Kall, P. (1986). “Approximation to Optimization Problems: An Elementary Review”. In: *Mathematics of Operations Research* 11.1, pp. 9–18. ISSN: 0364765X, 15265471. URL: http://www.jstor.org/stable/3690048.

Kato, K. (2012). “Weighted Nadaraya-Watson Estimation of Conditional Expected Shortfall”. In: *Journal of Financial Econometrics* 10.2, pp. 265–291. DOI: 10.1093/jjfinec/nbs002. eprint: http://jfec.oxfordjournals.org/content/10/2/265.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/10/2/265.abstract.

Kiefer, J. (1967). "On Bahadur's representation of sample quantiles." In: *Ann. Math. Stat.* 38, pp. 1323–1342. ISSN: 0003-4851. DOI: 10.1214/aoms/1177698690.

Knight, K. (1998). "Bootstrapping sample quantiles in non-regular cases." In: *Stat. Probab. Lett.* 37.3, pp. 259–267. ISSN: 0167-7152. DOI: 10.1016/S0167-7152(97)00125-9.

Knight, K. (2002). "What are the Limiting Distributions of Quantile Estimators?" In: *Statistical Data Analysis Based on the L1-Norm and Related Methods.* Ed. by Y. Dodge. Birkhäuser Basel, pp. 47–65. ISBN: 978-3-0348-8201-9. DOI: 10.1007/978-3-0348-8201-9_5.

Koenker, R. (2005). *Quantile regression.* Cambridge: Cambridge University Press, pp. xv +349. ISBN: 0-521-84573-4.

Kuan, C.-M., J.-H. Yeh, and Y.-Ch. Hsu (2009). "Assessing value at risk with CARE, the Conditional Autoregressive Expectile models". In: *Journal of Econometrics* 150.2. Recent Development in Financial Econometrics, pp. 261–270. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2008.12.002. URL: http://www.sciencedirect.com/science/article/pii/S0304407608002236.

Kusuoka, S. (2001). "On law invariant coherent risk measures". In: *Advances in Mathematical Economics.* Ed. by T. Kusuoka S. and Maruyama. Tokyo: Springer Japan, pp. 83–95. ISBN: 978-4-431-67891-5. DOI: 10.1007/978-4-431-67891-5_4. URL: https://doi.org/10.1007/978-4-431-67891-5_4.

Lambert, N.S. (2013). "Elicitation and Evaluation of Statistical Forecasts". unpublished. unpublished.

Lambert, N.S., D.M. Pennock, and Y. Shoham (2008). "Eliciting Properties of Probability Distributions". In: *Proceedings of the 9th ACM Conference on Electronic Commerce.* EC '08. Chicago, Il, USA: ACM, pp. 129–138. ISBN: 978-1-60558-169-9. DOI: 10.1145/1386790.1386813. URL: http://doi.acm.org/10.1145/1386790.1386813.

Lambert, N.S. and Y. Shoham (2009). "Eliciting Truthful Answers to Multiple-choice Questions". In: *Proceedings of the 10th ACM Conference on Electronic Commerce.* EC '09. Stanford, California, USA: ACM, pp. 109–118. ISBN: 978-1-60558-458-4. DOI: 10.1145/1566374.1566391. URL: http://doi.acm.org/10.1145/1566374.1566391.

Lehmann, E.L. and G. Casella (2006). *Theory of Point Estimation.* Springer Texts in Statistics. Springer New York. ISBN: 9780387227283. URL: https://books.google.de/books?id=4f24CgAAQBAJ.

Linton, O. and Z. Xiao (Jan. 2013). "Estimation of and Inference about the Expected Shortfall for Time Series with Infinite Variance". In: *Econometric Theory* 29.4, pp. 771–807. DOI: 10.1017/S0266466612000692. URL: https://www.cambridge.org/core/article/estimation-of-and-inference-about-the-expected-shortfall-for-time-series-with-infinite-variance/40B747D8235D51978C58CA8F3E5E7F5D.

McNeil, A.J., R. Frey, and P. Embrechts (2015). *Quantitative risk management. Concepts, techniques and tools. Revised edition.* Revised edition. Princeton, NJ: Princeton University Press, pp. xix+699. ISBN: 978-0-691-16627-8/hbk.

Merlevède, F. and E. Peligrad M. and Rio (Dec. 2011). "A Bernstein type inequality and moderate deviations for weakly dependent sequences". In: *Probability Theory and Related Fields* 151.3, pp. 435–474. DOI: 10.1007/s00440-010-0304-9.

Molchanov, I. (2005). *Theory of Random Sets*. Probability and Its Applications. Springer London.

Mosco, U. (1969). "Convergence of convex sets and of solutions of variational inequalities". In: *Advances in Mathematics* 3.4, pp. 510–585. ISSN: 0001-8708. DOI: https://doi.org/10.1016/0001-8708(69)90009-7. URL: http://www.sciencedirect.com/science/article/pii/0001870869900097.

Munkres, J.R. (2000). *Topology*. Featured Titles for Topology Series. Prentice Hall, Incorporated. ISBN: 9780131816299.

Neuhaus, G. (Aug. 1971). "On Weak Convergence of Stochastic Processes with Multidimensional Time Parameter". In: *Ann. Math. Statist.* 42.4, pp. 1285–1295. DOI: 10.1214/aoms/1177693241. URL: https://doi.org/10.1214/aoms/1177693241.

Newey, W.K. and D. McFadden (1994). "Chapter 36 Large sample estimation and hypothesis testing". In: *Handbook of Econometrics* 4, pp. 2111–2245. DOI: http://dx.doi.org/10.1016/S1573-4412(05)80005-4.

Newey, W.K. and J. Powell (1987). "Asymmetric least squares estimation and testing." In: *Econometrica* 55.4, pp. 819–847.

Nolde, N. and J.F. Ziegel (Dec. 2017). "Elicitability and backtesting: Perspectives for banking regulation". In: *Ann. Appl. Stat.* 11.4, pp. 1833–1874. DOI: 10.1214/17-AOAS1041. URL: https://doi.org/10.1214/17-AOAS1041.

Ogura, Y. (2007). "On some metrics compatible with the Fell–Matheron topology". In: *International Journal of Approximate Reasoning* 46.1. Special Section: Random Sets and Imprecise Probabilities (Issues in Imprecise Probability), pp. 65–73. ISSN: 0888-613X. DOI: https://doi.org/10.1016/j.ijar.2006.12.008. URL: http://www.sciencedirect.com/science/article/pii/S0888613X06001630.

Oliveira, P.E. and Ch. Suquet (1998). "Weak convergence in Lp(0,1) of the uniform empirical process under dependence". In: *Statistics & Probability Letters* 39.4, pp. 363–370. ISSN: 0167-7152. DOI: https://doi.org/10.1016/S0167-7152(98)00091-1.

Osband, K. (1985). "Providing Incentives for Better Cost Forecasting". Ph. D. Thesis.

Patton, A.J. (Mar. 2016). *Evaluating and Comparing Possibly Misspecified Forecasts*. Working Paper. Department of Economics, Duke University. URL: https://public.econ.duke.edu/~ap172/Patton_bregman_comparison_27mar15.pdf.

Peracchi, F. and A.V. Tanase (2008). "On estimating the conditional expected shortfall". In: *Applied Stochastic Models in Business and Industry* 24.5, pp. 471–493. ISSN: 1526–4025. DOI: 10.1002/asmb.729. URL: http://dx.doi.org/10.1002/asmb.729.

Pomarede, J. (1976). "unified approach via graphs to Skorohod's topologies on the function space D". Ph.D. thesis.

Rio, E. (Sept. 2013). "Inequalities and limit theorems for weakly dependent sequences". 3rd cycle. Lecture.

Rockafellar, R.T. and Stanislav Uryasev (2002). "Conditional value-at-risk for general loss distributions". In: *Journal of Banking & Finance* 26.7, pp. 1443–1471. ISSN: 0378-4266. DOI: https://doi.org/10.1016/S0378-4266(02)00271-6. URL: http://www.sciencedirect.com/science/article/pii/S0378426602002716.

Savage, L.J. (1971). "Elicitation of Personal Probabilities and Expectations". In: *Journal of the American Statistical Association* 66.336, pp. 783–801. ISSN: 01621459. URL: http://www.jstor.org/stable/2284229.

Scaillet, O. (2004). "Nonparametric Estimation and Sensitivity Analysis of Expected Shortfall". In: *Mathematical Finance* 14.1, pp. 115–129. ISSN: 1467-9965. DOI: 10.1111/j.0960-1627.2004.00184.x. URL: http://dx.doi.org/10.1111/j.0960-1627.2004.00184.x.

Skorokhod, A.V. (1956). "Limit Theorems for Stochastic Processes". In: *Theory of Probability & Its Applications* 1.3, pp. 261–290. DOI: 10.1137/1101022. eprint: https://doi.org/10.1137/1101022. URL: https://doi.org/10.1137/1101022.

Smirnov, N.V. (1952). "Limit distribution for the terms of a variational series: (Trudy matematičeskogo instituta imeni V. A. steklova 25, 60 pp. (1949))". In: *American Mathematical Society Translations* 11.1, pp. 82–143.

Steinwart, I., C. Pasin, R. Williamson, and S. Zhang (June 2014). "Elicitation and Identification of Properties". In: *Proceedings of The 27th Conference on Learning Theory*. Ed. by M.F. Balcan, V. Feldman, and C. Szepesvári. Vol. 35. Proceedings of Machine Learning Research. Barcelona, Spain: PMLR, pp. 482–526. URL: http://proceedings.mlr.press/v35/steinwart14.html.

Straf, M.L. (1969). *A General Skorohod Space and Its Application to the Weak Convergence of Stochastic Processes with Several Parameters*. University of Chicago, Department of Statistics. URL: https://books.google.de/books?id=mgKtSgAACAAJ.

Taylor, J.W. (2008). "Estimating Value at Risk and Expected Shortfall Using Expectiles". In: *Journal of Financial Econometrics* 6.2, pp. 231–252. DOI: 10.1093/jjfinec/nbn001. eprint: http://jfec.oxfordjournals.org/content/6/2/231.full.pdf+html. URL: http://jfec.oxfordjournals.org/content/6/2/231.abstract.

Thomson, B.S., A.M. Bruckner, and J.B. Bruckner (2008). *Real Analysis*. www.classicalrealanalysis.com.

van der Vaart, A.W. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 9780521784504.

van der Vaart, A.W. and Jon A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer. ISBN: 9780387946405.

Weber, S. (2006). "Distribution-Invariant Risk Measures, Information, and Dynamic Consistency". In: *Mathematical Finance* 16.2, pp. 419–441. DOI: 10.1111/j.1467-9965.2006.00277.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9965.2006.00277.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9965.2006.00277.x.

Whitt, W. (2002). *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues*. Springer Series in Operations Research and Financial Engineering. Springer New York.

Wijsman, R. A (Jan. 1964). “Convergence of Sequences of Convex Sets, Cones and Functions”. In: *Bull. Amer. Math. Soc.* 70.1, pp. 186–188. URL: `https://projecteuclid.org:443/euclid.bams/1183525799`.

Wijsman, R. A. (1966). “Convergence of Sequences of Convex Sets, Cones and Functions. II”. In: *Transactions of the American Mathematical Society* 123.1, pp. 32–45. ISSN: 00029947. URL: `http://www.jstor.org/stable/1994611`.

Ziegel, J.F. (2016). “Coherence and elicitability”. In: *Math. Finance* 26.4, pp. 901–918. ISSN: 0960-1627; 1467-9965/e. DOI: `10.1111/mafi.12080`.

Ziegel, J.F., F. Krüger, A. Jordan, and F. Fasciati (2017). *Murphy Diagrams: Forecast Evaluation of Expected Shortfall*. Discussion Paper Series 632. urn:nbn:de:bsz:16-heidok-229502. Heidelberg: University of Heidelberg, Department of Economics. DOI: `10.11588/heidok.00022950`. URL: `http://hdl.handle.net/10419/162975`.

Erklärung

Hiermit versichere ich, dass ich die vorliegende Dissertation

Asymptotics for selected Risk Measures under general assumptions

selbst und ohne fremde Hilfe verfasst habe. Ich habe keine weiteren Quellen oder Hilfsmittel benutzt als angegeben und habe vollständige oder sinngemäße Zitate als solche gekennzeichnet.

Dies ist mein erster Versuch einer Promotion. Die Dissertation wurde bisher weder in der vorliegenden, noch in ähnlicher Form bei einer anderen in- oder ausländischen Hochschule anlässlich eines Promotionsgesuchs oder zu anderen Prüfungszwecken eingereicht.

Tobias Zwingmann
Marburg, 1. November 2018

Zusammenfassung

In der vorliegenden Arbeit beschäftigen wir uns mit der Analyse von asymptotischen Eigenschaften ausgewählter Risikomaße unter allgemeinen Annahmen.

Betrachten wir einen Mechanismus in der realen Welt, sollten Aussagen über dessen zukünftigen Zustand von probabilistischer Natur sein, sodass Abweichungen der Vorhersage von dem später beobachteten Zustand möglich sind. Diese Abweichungen werden als *Risiko* bezeichnet. *Risikomaße* bieten eine Möglichkeit, dieses Risiko zu quantifizieren, sodass beispielsweise Entscheidungen auf Grundlage dieser Werte möglich werden. Normalerweise müssen die Risikomaße in der Praxis geschätzt werden, da der „wahre" Mechanismus unbekannt ist.

Um Risikomaße vernünftig anwenden zu können, sollten sie gewisse Eigenschaften erfüllen, die allgemein anerkannte Prinzipien von „Risiko", insbesondere Prinzipien aus der Finanzmathematik, widerspiegeln. Dazu gehören:

i) *Normalisierung*: Wir sollten keinem Risiko ausgesetzt sein, wenn wir keine Anlage besitzen;

ii) *Translationsinvarianz*: Fügen wir eine Anlage mit sicherer Rendite zu unserem Portfolio hinzu, sollte sich das Risiko des gesamten Portfolios um diesen Betrag verringern;

iii) *Monotonie*: Hat eine Anlage immer eine bessere Rendite als eine andere, sollte erstere ein höheres Risiko besitzen;

iv) *Sub-Additivität*: Das Risiko eines Portfolios darf die Summe der Risiken der einzelnen Positionen nicht überschreiten (*Risikodiversifizierung*);

v) *Positive Homogenität*: Kaufen wir einen anderen Anteil einer Anlage, sollte das Risiko entsprechend skalieren.

Risikomaße, die diese Eigenschaften erfüllen, werden *kohärent* genannt und sind wichtige Kriterien für die Wahl eines Risikomaßes.

Das erste Maß, mit dem wir uns während der Arbeit beschäftigen, ist der *Value at Risk*. Dieser ist eines der ältesten genutzten Risikomaße und wird vom Basler Ausschuss für Bankenaufsicht als zu nutzendes Risikomaß vorgeschlagen (Basel, 2017; Basel, 2013).

Definitionsgemäß ist der Value at Risk zu einem fixierten Level $\alpha \in (0,1)$ ein Quantil der Verlustverteilung F und beantwortet somit die Frage, welcher Verlust mit einer Wahrscheinlichkeit größer als (oder gleich) α nicht überschritten wird. Alternativ kann der Value at Risk als Minimum einer Kontrastfunktion definiert werden, was viele statistische Methoden wie Regression (Koenker, 2005) und Simulation auf Basis historischer Werte ermöglicht („comparative Backtests“). In der Literatur werden viele Schätzer für Quantile vorgeschlagen und auf das asymptotische Verhalten – wie Konsistenz für das wahre Quantil und schwache Konvergenz – untersucht. Wir wählen hauptsächlich das *empirische Quantil* $\widehat{q}_{n,\alpha}$ einer Stichprobe der Größe n als Schätzer für das wahre Quantil q_α, für welches die schwache Konvergenz unter Standardannahmen hinreichend untersucht wurde. Genauer gilt, dass, wenn die Verlustfunktion F im betrachteten Quantil differenzierbar mit positiver Ableitung ist, die Folge $\sqrt{n}(\widehat{q}_{n,\alpha} - q_\alpha)$ schwach gegen eine Normalverteilung konvergiert.

Problematisch bei der Anwendung des Value at Risk ist die fehlende sub-Additivität, sodass Diversifikation verhindert werden könnte. Als Alternative wurde daher der *Expected Shortfall* eingeführt, welches das zweite Risikomaß ist, das wir in der Arbeit betrachten. Es ist definiert als Mittel der Value at Risk Werte, die bis zu einem fixierten Level auftreten, und beantwortet somit die Frage, wie hoch der Verlust im Mittel über die schlechtesten $\alpha\,100\%$ der Fälle ist. Mittlerweile wird auch der Expected Shortfall vom Basler Ausschuss für Bankenaufsicht zur Nutzung vorgeschrieben (Basel, 2013).

Es stellt sich heraus, dass der Expected Shortfall ein kohärentes Risikomaß ist, leider aber nicht als Minimum einer geeigneten Kontrastfunktion in *einer* Variablen definiert werden kann. Ersteres unterstützt die Wahl des Expected Shortfalls als Risikomaß, letzteres macht eine Anwendung fraglich. Fissler und Ziegel (2016) konnten aber zeigen, dass das Paar (Value at Risk, Expected Shortfall) das Minimum einer Kontrastfunktion in *zwei* Variablen ist.

Auch für den Expected Shortfall wurden verschiedene Schätzer vorgeschlagen und auf das asymptotische Verhalten untersucht. Wir betrachten den Schätzer $\widehat{es}_{n,\alpha}$, welcher als Minimierer der empirischen Kontrastfunktion aus Fissler und Ziegel (2016) entsteht. Unter den Regularitätsannahmen an F wie oben, besitzt auch $\sqrt{n}(\widehat{es}_{n,\alpha} - es_\alpha)$ eine Normalverteilung als schwachen Grenzwert.

In der Arbeit befinden wir uns in der folgenden Situation: Ist die Verlustfunktion F im gewählten α-Quantil nicht regulär, können schwache Grenzwerte der Folge $a_n(\widehat{q}_{n,\alpha} - q_\alpha)$ auftreten, die nicht-normal sind (Knight, 2002). Darüber hinaus muss in diesem Fall oft eine Konvergenzrate a_n gewählt werden, welche $\frac{\sqrt{n}}{a_n} \to \infty$ erfüllt. Betrachten wir nun in einer solchen Situation den bivariaten Parameter (Value at Risk, Expected Shortfall), stellt sich die Frage, ob auch für den Expected Shortfall eine andere Konvergenzrate gewählt werden muss und ob sich der schwache Grenzwert ändert. Wie wir gezeigt haben, ist das nicht der Fall: Das Konvergenzverhalten des Quantils hat keinen Einfluss auf die Konvergenzrate oder die asymptotische Verteilung des Expected Shortfalls.

Unser Resultat formulieren wir auch für den multivariaten Fall, in dem mehrere Level $\alpha_1, \ldots, \alpha_s$ gleichzeitig betrachtet werden. Außerdem verallgemeinern wir das erzielte Ergebnis auf eine größere Klasse von Risikomaßen, genauer auf solche, deren erster Eintrag ein Bayes-Schätzer und deren zweiter Eintrag das zugehörige Bayes-Risiko ist. Diese Klasse beinhaltet das Paar (Value at Risk, Expected Shortfall) als Spezialfall.

Ein drittes, weitverbreitetes Risikomaß ist das *Expektil* (Newey und Powell, 1987). Das Expektil μ_τ kann als Minimum einer Kontrastfunktion definiert werden und ist kohärent; dies ist eine einmalige Eigenschaft unter Risikomaßen (Ziegel, 2016). Während der Expected Shortfall nur Verluste über einem gewissen Level gewichtet, berücksichtigt das Expektil auch Abweichungen nach unten; eine wirtschaftliche Rechtfertigung dafür ist in Ehm u. a. (2016) zu finden.

Wie für die bisherigen Risikomaße betrachten wir das empirische Analogon des Expektils, also den Schätzer $\widehat{\mu}_{\tau,n}$ welcher als Minimierer der empirischen Kontrastfunktion gegeben ist. Betrachten wir die Folge von Abbildungen $\tau \mapsto \sqrt{n}(\widehat{\mu}_{\tau,n} - \mu_\tau)$, $\tau \in [\tau_l, \tau_u] \subset (0,1)$, so kann dies als stochastischer Prozess (mit fast sicher stetigen Pfaden) interpretiert werden, dieser trägt den Namen *empirischer Expektil Prozess*. Für dieses Objekt haben Holzmann und Klar (2016) schwache Konvergenz bezogen auf die Supremums-Norm gegen einen Gauß'schen Prozess gezeigt, falls die zugrunde liegende Verteilung F stetig auf einer Umgebung von $[\tau_l, \tau_u]$ ist. Ebenso wurde gezeigt, dass ein nicht-normaler Grenzwert der Folge $\sqrt{n}(\widehat{\mu}_{\tau_0,n} - \mu_{\tau_0})$ auftritt, falls F in μ_{τ_0} Masse besitzt. Wir untersuchen in diesem Fall die schwache Konvergenz des empirischen Expektil Prozesses, wobei dann der Grenzprozess keine stetigen Pfade mehr besitzen kann. Damit ist die Supremum-Norm für diese Situation nicht geeignet, da in dieser eine stetige Funktion nicht gegen eine unstetige Funktion konvergieren kann. Wir nutzen daher die hypi-Semimetrik, welche von Bücher u. a. (2014) für solche Situationen vorgeschlagen wurde.

Zuletzt betrachten wir die Verteilungskonvergenz des empirischen Quantil Prozesses $\alpha \mapsto \sqrt{n}(\widehat{q}_{n,\alpha} - q_\alpha)$ unter schwachen Annahmen an die Verteilung F. Auch dieses Resultat verallgemeinert bekannte Ergebnisse über das Quantil (van der Vaart und Wellner, 1996, Example 3.9.24).

Die Arbeit ist wie folgt aufgebaut. Kapitel 1 definiert allgemeine Risikomaße, führt die oben beschriebenen Risikomaße rigoros ein und liefert weitere Zusammenhänge. Ein kurzer Überblick über Kontrastfunktionen (beziehungsweise *Score Funktionen*) wird ebenfalls gegeben, wo wir einen Beitrag zu aktuellen Fragestellungen über *Zerlegungen in elementare Score Funktionen* leisten (Diebold und Mariano, 1995; Clark und McCracken, 2001; Ehm u. a., 2016; Ziegel u. a., 2017). Dann wenden wir uns der Theorie der M-Schätzung zu; dort verallgemeinern wir ein bekanntes Resultat über Konvergenzraten von M-Schätzern (van der Vaart, 1998, Theorem 5.52) auf multivariate Schätzer, in denen die einzelnen Einträge unterschiedliche Konvergenzraten besitzen können.

Kapitel 2 wiederholt die schwache Konvergenztheorie für Prozesse mit Werten in metrischen Räumen nach Hoffmann-Jørgensen und zeigt, wie diese auf semimetrische Räume übertragen werden kann (Bücher u. a., 2014). Weiter führen wir die *Klammer-*

Entropie von Funktionenklassen ein, welche wir im Verlauf der Arbeit nutzen, um die schwache Konvergenz von empirischen Prozessen nachzuweisen. Nach diesem Einblick wenden wir uns konkreten Topologien zu, die auf dem Raum der beschränkten Funktionen beziehungsweise auf dem Raum der rechtsstetigen Funktionen mit linksseitig existierendem Grenzwert definiert werden. Zum einen führen wir die hypi-Topologie nach Bücher u. a. (2014) ein, die auf der Konvergenz von Mengen, welche die Funktionen charakterisieren, basiert. In unseren Anwendungen ist diese Topologie äquivalent zur Skorohod M_2-Topologie, die wir daher ebenfalls, zusammen mit der verwandten M_1-Topologie, erläutern. In diesem Abschnitt beweisen wir ein abstraktes Resultat über die schwache Konvergenz von Prozessen in der hypi-Topologie, was im Fall der empirischen Expektil und Quantil Prozesse angewendet werden kann, und zeigen die Äquivalenz der genannten Topologien.

In Kapitel 3 beweisen wir die beschriebenen Aussagen über das Risikomaß (Value at Risk, Expected Shortfall) und Erweiterungen davon. Wir betrachten zunächst das Aussehen der Schätzer und geben die Konvergenzraten an. Daraufhin bestimmen wir die asymptotische Verteilung der Schätzer und illustrieren das Konvergenzverhalten mit einer kurzen Simulation.

Die Ideen aus Kapitel 3 erweitern wir in Kapitel 4 auf Risikomaße bestehend aus einem Bayes-Schätzer und dem zugehörigen Bayes-Risiko.

Abschließend behandeln wir in Kapitel 5 die schwache Konvergenz der empirischen Expektil und Quantil Prozesse in der hypi-Topologie unter schwachen Annahmen an die zugrunde liegende Verteilung F. Im Falle des Expektils benötigen wir nur endliche zweite Momente der Beobachtungen; für das Quantil benötigen wir strikte Monotonie von F und eine Version der (fast überall definierten) Ableitung f von F, die von Null weg beschränkt ist und überall rechts- und linksseitige Grenzwerte besitzt. Die Ergebnisse für den Expektil Prozess werden numerisch illustriert.

Literatur

Basel (2013). *Fundamental review of the trading book: A revised market risk framework. https://www.bis.org/publ/bcbs265.pdf.*

Basel (2017). *Basel III: A global regulatory framework for more resilient banks and banking systems. https://www.bis.org/publ/bcbs189.pdf.*

Bücher, A., J. Segers und S. Volgushev (2014). „When uniform weak convergence fails: empirical processes for dependence functions and residuals via epi- and hypographs.“ In: *Ann. Stat.* 42.4, S. 1598–1634. ISSN: 0090-5364; 2168-8966/e. DOI: 10.1214/14-AOS1237.

Clark, T.E. und M.W. McCracken (2001). „Tests of equal forecast accuracy and encompassing for nested models“. In: *Journal of Econometrics* 105.1. Forecasting and empirical methods in finance and macroeconomics, S. 85–110. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(01)00071-9. URL: http://www.sciencedirect.com/science/article/pii/S0304407601000719.

Diebold, F.X. und R.S. Mariano (1995). „Comparing Predictive Accuracy“. In: *Journal of Business & Economic Statistics* 13.3, S. 253–265. DOI: 10.1198/073500102753410444.

Ehm, W., T. Gneiting, A. Jordan und F. Krüger (2016). „Of quantiles and expectiles: consistent scoring functions, Choquet representations and forecast rankings“. In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 78.3, S. 505–562. ISSN: 1467-9868. DOI: 10.1111/rssb.12154. URL: http://dx.doi.org/10.1111/rssb.12154.

Fissler, T. und J.F. Ziegel (Aug. 2016). „Higher order elicitability and Osband's principle“. In: *Ann. Statist.* 44.4, S. 1680–1707. DOI: 10.1214/16-AOS1439. URL: https://doi.org/10.1214/16-AOS1439.

Holzmann, H. und B. Klar (2016). „Expectile asymptotics“. In: *Electron. J. Statist.* 10.2, S. 2355–2371. DOI: 10.1214/16-EJS1173. URL: http://dx.doi.org/10.1214/16-EJS1173.

Knight, K. (2002). „What are the Limiting Distributions of Quantile Estimators?“ In: *Statistical Data Analysis Based on the L1-Norm and Related Methods.* Hrsg. von Y. Dodge. Birkhäuser Basel, S. 47–65. ISBN: 978-3-0348-8201-9. DOI: 10.1007/978-3-0348-8201-9_5.

Koenker, R. (2005). *Quantile regression.* Cambridge: Cambridge University Press, S. xv +349. ISBN: 0-521-84573-4.

Newey, W.K. und J. Powell (1987). „Asymmetric least squares estimation and testing.“ In: *Econometrica* 55.4, S. 819–847.

van der Vaart, A.W. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press. ISBN: 9780521784504.

van der Vaart, A.W. und Jon A. Wellner (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* Springer Series in Statistics. Springer. ISBN: 9780387946405.

Ziegel, J.F. (2016). „Coherence and elicitability“. In: *Math. Finance* 26.4, S. 901–918. ISSN: 0960-1627; 1467-9965/e. DOI: `10.1111/mafi.12080`.

Ziegel, J.F., F. Krüger, A. Jordan und F. Fasciati (2017). *Murphy Diagrams: Forecast Evaluation of Expected Shortfall.* Discussion Paper Series 632. urn:nbn:de:bsz:16-heidok-229502. Heidelberg: University of Heidelberg, Department of Economics. DOI: `10.11588/heidok.00022950`. URL: `http://hdl.handle.net/10419/162975`.

Curriculum Vitæ